

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2004, 22:08:49 ; Search time 3667.2 Seconds
(without alignments)
10389.001 Million cell updates/sec

Title: US-10-016-604-5

Perfect score: 879

Sequence: 1 tggggggaagaagaagaagag.....ccctctctctctctctctata 879

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6340544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl1:
1: gb_da:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_un:*
28: em_vl:*
29: em_vl:*
30: em_hg_hum:*
31: em_hg_inv:*
32: em_hg_other:*
33: em_hg_mus:*
34: em_hg_pln:*
35: em_hg_rtd:*
36: em_hg_mam:*
37: em_hg_vrt:*
38: em_hg_hum:*
39: em_hg_mus:*
40: em_hg_hum:*
41: em_hg_hum:*
42: em_hg_hum:*
43: em_hg_hum:*
44: em_hg_hum:*
45: em_hg_hum:*

Pred. No. 1e the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	879	100.0	879	AX593927	AX593927 Sequence
2	879	100.0	968	AX594072	AX594072 Sequence
3	879	100.0	9472	AF164614	AF164614 Homo sapi
4	879	100.0	19221	AF074086	AF074086 Homo sapi
5	877.4	99.8	154037	AC072054	AC072054 Homo sapi
6	876	99.7	8300	HEN17834	Y17834 Human endog
7	875.8	99.6	8708	AF490464	AF490464 Homo sapi
8	874.2	99.5	10029	HEN17832	Y17832 Human endog
9	871	99.1	6206	AF261945	AF261945 Homo sapi
10	870	99.0	1505	HERVCONF	X82271 Human endog
11	866.4	98.6	3366	HERVCONF	X82272 Human endog
12	864.6	98.4	94145	AC025420	AC025420 Homo sapi
13	864.6	98.4	181915	AC110927	AC110927 Homo sapi
14	864.6	98.4	205317	AC093856	AC093856 Homo sapi
15	861.4	98.0	122913	AL139404	AL139404 Human DNA
16	861.4	98.0	152052	AC016728	AC016728 Homo sapi
17	861.4	98.0	154286	AC016698	AC016698 Homo sapi
18	861.4	98.0	166460	AC006366	AC006366 Homo sapi
19	861.4	98.0	168843	AC068207	AC068207 Homo sapi
20	861.4	98.0	176257	AC006039	AC006039 Homo sapi
21	861.4	98.0	185588	AP006305	AP006305 Homo sapi
22	861.4	98.0	188712	AC016582	AC016582 Homo sapi
23	861.4	98.0	197245	AL591164	AL591164 Homo sapi
24	861.4	98.0	199654	AC010898	AC010898 Homo sapi
25	859.8	97.8	9181	AF164611	AF164611 Homo sapi
26	859.8	97.8	112684	AC106051	AC106051 Homo sapi
27	859.8	97.8	161562	AC024626	AC024626 Homo sapi
28	859.8	97.8	180634	AC012175	AC012175 Homo sapi
29	859.8	97.8	182372	AC006035	AC006035 Homo sapi
30	859.8	97.8	183976	AC100821	AC100821 Homo sapi
31	858.2	97.6	163455	AC009427	AC009427 Homo sapi
32	858.2	97.6	164168	AL135927	AL135927 Human DNA
33	858.2	97.6	164179	AC007227	AC007227 Homo sapi
34	858.2	97.6	186294	AC021108	AC021108 Homo sapi
35	858.2	97.6	195880	AC074117	AC074117 Homo sapi
36	858.2	97.6	216342	AC009132	AC009132 Homo sapi
37	858.2	97.6	227611	AC023766	AC023766 Homo sapi
38	858.2	97.6	322058	AC068887	AC068887 Homo sapi
39	856.6	97.5	9463	AY037929	AY037929 Human end
40	856.6	97.5	15000	AX800318	AX800318 Sequence
41	856.6	97.5	15000	AX800318	AX800318 Sequence
42	856.6	97.5	69676	AL596188	AL596188 Human DNA
43	856.6	97.5	98954	HS65708	AL034407 Human DNA
44	856.6	97.5	121914	AC124944	AC124944 Homo sapi
45	856.6	97.5	131011	AC007390	AC007390 Homo sapi

ALIGNMENTS

RESULT 1
AX593927
LOCUS AX593927 879 bp DNA linear PAT 13-FEB-2003
DEFINITION Sequence 5 from Patent WO0246477.
ACCESSION AX593927
VERSION AX593927.1 GI:28375186
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Garcia, P., Hardy, S.F., Williams, L.T. and Escobedo, J.
TITLE Endogenous retroviruses up-regulated in prostate cancer
JOURNAL Patent: WO 0246477-A 5 13-JUN-2002;

Db 661 CACGACCTTAATCTTTACATGTCATGATGCAAGACCTTTGTCACATGTTGTCT 720
Qy 721 GGTGACCTCTCCACACATTTGTTGACCCCTGACACATCCCTCTTGAGAAACAC 780
Db 721 GGTGACCTCTCCACACATTTGTTGACCCCTGACACATCCCTCTTGAGAAACAC 780
Qy 781 CCACAGATGATGATTAATACTAAGGAACTCAGAGCTGGCGGATCTTCATATCTG 840
Db 781 CCACAGATGATGATTAATACTAAGGAACTCAGAGCTGGCGGATCTTCATATCTG 840
Qy 841 AAGCTGCTTCCCGGCTCCCTCTTTCTTTCTCTATA 879
Db 841 AAGCTGCTTCCCGGCTCCCTCTTTCTTTCTCTATA 879
RESULT 3
AF164614 9472 bp DNA linear PRI 03-SEP-1999
LOCUS Homo sapiens endogenous retrovirus HERV-K108, complete sequence.
DEFINITION AF164614
ACCESSION AF164614 GI:5802820
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 9472)
AUTHORS Bariltescu,M., Turner,G., Seaman,M.I., Delnard,A.S., Kidd,K.K. and Lenz,J.
TITLE Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans
JOURNAL (Curt. Biol. 9 (16), 861-868 (1999)
MEDLINE 10469592
PUBMED 2 (bases 1 to 9472)
REFERENCE Bariltescu,M., Turner,G., Seaman,M.I., Delnard,A.S., Kidd,K.K. and Lenz,J.
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (02-JUL-1999) Molecular Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA
FEATURES
source Location/Qualifiers
1. 9472
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1. 9472
/rpt_family="human endogenous retrovirus HERV-K108"
1. 968
1112..6749
/gene="gag-pro-pol"
/gene="gag-pro-pol"
/join(1112..3108,3108..3915,3915..6749)
/note="-1 frameshifts presumed to occur within the codons for the last amino acids in the Gag and Pro open reading frames"
/codon_start=1
/product="Gag-Pro-Pol protein"
/protein_id="A051797.1"
/db_xref="GI:5802821"
/translation="MGOTSKIKSVSYLSFIKILKRGVKTSTNLIKLFQIIBO
FCMPFPGOTLIDKDMRIKELKOKARKNIIPITVMWMAIKALFPPOMEDSV
SVSDAPSCITIDCENRTKRSOKETBEIHEVNAEPMASTONVDVNOQOEVYPT
LKLBGKPELVGPEBESKPKGTSPLACQVPTLLOPKOVENKTOPVAOYMPFAL
QYRPPESQYQYGPMPAPQGRAPYPOPPRRILNTPAPPSRQSKLHEI1DKSKED
TEAWQFVTLLEPMPEGAGQGEPTVEARYSFSIKLDMKESGVQYSPNSYMT
LDSIAHGRLIPYDMEILAKSLSLSPQLOFTWIDGOVQVRRAANPUNIDA
DOLGIGONSTISQALMONEATEIYORALICRAMEKIDQSGTSPFNTRQSGKEP
YPPVARLOVNAKSIADBEKARKYIVLEAMENAPRCQAKIKELKVVAGSVISE
YVACDGTIGAMKAMLAQNTIGVIGGVRTGRKCTNCGQIHLKACQVILKON
IT1QATYGRBPDLPCLCRCKGKRWASQCSKFKNGQPSGNORCPQAPQDTGAF
PIQFVPOGPOGQPLSOVFOGISOPLPOVNNCPPOAAYOVLDCTIOAVSLGEP
PQKPTGVGVLPGKVTGLIGRSILNKGVOITSTVSDDYKGEIOLVSSISIPMA
SPDRIRIAQLLLPFIYKGNSEIKRIGIGSGTDPGKAAYASQYSENRPYCKAIIOGK

QFEGVDTGADVSIILNOMPKNPKQKAVTGLVIGTASVEVYOSTEILHCLBPNQ
STVQPMITTSIPILNLMGRDLQOWGAETWPAPEPSPIQKIMTGMGIPKQKEND
GKIYPAKINORNEGIDNLCGAATEPEKPIPEITKTEKPAVWNOMPEKQKEL
HILANOLEKGIHEBSPSPNSPVFVIOKSGKRMITDIAVAVIOPMGQGLP
SPAMIPKQWPLI1IDDKCFPTPLABODEKRAFTIPALNNKPAIRFOKUTPOG
LNSPTICQYTVGRALQVREKFSDCYIHCIDIDILCAETVDKIDCYTHQAVNA
GLIASDKIOTSPPFHYLMQJENRKKPKPIH1KQVETLNDFOQLADIMIRPT
LGIPTAMSNLFSILRGSDNLSKRMTPKATKEIKLVEKIQSOANR1IDPLAPL
LIPATASPTGII1IONDLYMSFSTVCSKAFKPLDQKISHTGIPYNSQOAV
PDKIYVPLTEQVROAFINSKAMKIGIANVGI1DNNYPRKTIPOKILTWILKGT
RREPLENALTVFTDSSNGKAAATGPERIKTPYQSAQAEVAVTVUODBPQIN
I1SDSAVQVATDVELTALIKYSMDQDNLQVFLQVTRKRNPEYTHIRATNCP
GPLTKNEQADLVSSALIKAOEILHALTHVNAAGLNKRPVTKQAKD1VOHQOQV
LHLPTQAGVNPGLCPNALMOMDVTVPSFGLSYVHVTVDYSHF1MAQCQEST
SHVKHLSQCFVAVGVPKIKTDNGPBGCSKAFKPLDQKISHTGIPYNSQOAV
ERTNRTILKQVQKQEGDSKECTTPQMOALALYTLNFINIYRNQTTSAEOLNCK
KNSPHEGKLI1WPKDNKKTVEIGVITVWGFVFCVSGENQLPVMTTRHUKTNEI
RDAPKSTASLETTSOSTVSDQONQNDVARTBVALHQBGAANIGTTRKADAVSYK
ISREHKDITPREVAAACSLDCTINGKSPYACRS5CS"
6451..8550
/gene="env"
6451..8550
/gene="env"
/codon_start=1
/product="envelope protein"
/protein_id="A051798.1"
/db_xref="GI:5802822"
/translation="MNPSEMQKAPRRRRHRRAPLTHKQNKVTSSEOMKLPSTK
AAPPVMAQLKLQIOLATKYLENTRKVOTPSWMLIDRCAKEEBSMMN1SGHY
YMAVPPPIRAVATWMDNTEVYVNDSPVWPBIDRCAKEEBSMMN1SGHY
PPIICGAPGCLMPAVONMLVEVTPVSPICRFTYHWSGMSLPRVYLODFORS
KFPKPKGPKKEI1PKSKNTEV1VWECVANSAYILONNEFTI1DVAQGFYHNS
GOROSGSAQVSPAVSDLTRESLDKHKHKLQSPYMEKKEKISTPRKTVSGSP
EHPBLRILTAASHIRIRWSGNQTLERDRKPFITIDNSSLTPLQSCYKPRVLYVG
NIVIKPDSOTTCENCLTICIDSTFWQRILVLRBEVW1PVSQDRWESVSH
ILTEVLKGVNRSRPIFTL1AVIMGLIATADAAGVALHSVQSVNVDQKNS
TRLNQSOS1DOKLANIDRQVITVMSGRMSLEHFPOLQCDMNTSDCTPOILN
ESEHNDMVRH1OGREDNL1DLSIKKEQIFEASKALN1LVPQETM1AGVADLANL
NPYTWYTGSTTILN1LIVCLFCLLLVCRCTQQRSDRHRERAMTAVLSKRX
GNVGSKRPQIYTVSV"
8504..9472
ORIGIN
LTR
Query Match 100.0%; Score 879; DB 9; Length 9472;
Best Local Similarity 100.0%; Pred. No. 1,1e-262;
Matches 879; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTCTGTGTAGAAAAGTAGACAT 60
Db 1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTCTGTGTAGAAAAGTAGACAT 60
Qy 61 AGGAGACTCCATTTTGTATGACTAAGAAAATTTCTTGCTTGAGATTCTGTATC 120
Db 61 AGGAGACTCCATTTTGTATGACTAAGAAAATTTCTTGCTTGAGATTCTGTATC 120
Qy 61 AGGAGACTCCATTTTGTATGACTAAGAAAATTTCTTGCTTGAGATTCTGTATC 120
Db 61 AGGAGACTCCATTTTGTATGACTAAGAAAATTTCTTGCTTGAGATTCTGTATC 120
Qy 121 TATGACCTTACCCCAACCCCGCTCTGAAAACAGTGTGTGCACACTCAGGGTTAA 180
Db 121 TATGACCTTACCCCAACCCCGCTCTGAAAACAGTGTGTGCACACTCAGGGTTAA 180
Qy 181 ATGATTAAGGCGGTGACAGATGTGCTTTGTTAAAGATGCTTGAAGGACGATCTC 240
Db 181 ATGATTAAGGCGGTGACAGATGTGCTTTGTTAAAGATGCTTGAAGGACGATCTC 240
Qy 241 CTTAAGAGTATCAACAATCTCTTAATCTCAAGTACCAAGGACCAAAAATCTGGGAAAG 300
Db 241 CTTAAGAGTATCAACAATCTCTTAATCTCAAGTACCAAGGACCAAAAATCTGGGAAAG 300
Qy 301 CCGAAGGACCTCTGCTAGGAAACCAAGATTTGTCACAGTTTCCCAATGATAG 360
Db 301 CCGAAGGACCTCTGCTAGGAAACCAAGATTTGTCACAGTTTCCCAATGATAG 360
Qy 361 CTTGAAATATAGCTCTGTGGGAAAGGAAACCTGACCGTCCCAACCGGACACCGGTA 420
Db 361 CTTGAAATATAGCTCTGTGGGAAAGGAAACCTGACCGTCCCAACCGGACACCGGTA 420

Db	361	CCTGAAAATATGAGCCCTGTGGGAAAGGAAAGAACTGACCGTCCCGACGCCGACACCCGTA	420
Qy	421	AAAGGCTCTGTGCTAGAGAGATTAAGTAAAAAAGAAAGAAATGCCCTTGTGACGTTGAGCA	480
Db	421	AAAGGCTCTGTGCTAGAGAGATTAAGTAAAAAAGGAAAGAAATGCCCTTGTGACGTTGAGCA	480
Qy	481	AGAGAAAGCAATCTGTCTCTGACCTGTCCCTGGGCAATGGAATGTCGATATAAAACC	540
Db	481	AGAGAAAGCAATCTGTCTCTGACCTGTCCCTGGGCAATGGAATGTCGATATAAAACC	540
Qy	541	GATTGTATGCTCCATCTAATGAGATAGGAAAAAACCCGCTTAAGGGCTGAGGTGGACCT	600
Db	541	GATTGTATGCTCCATCTAATGAGATAGGAAAAAACCCGCTTAAGGGCTGAGGTGGACCT	600
Qy	601	CGGGGACCAATATCTGCTTTGTGAAAGCACTGAGATGTTATGTATGATGATCTAAAG	660
Db	601	CGGGGACCAATACTGCTTTGTGAAAGCACTGAGATGTTATGTATGATGATCTAAAG	660
Qy	661	CACAGCATTTATCCTTTACATATTCATATGACAAAGACCTTGTTCACATGTTGTCT	720
Db	661	CACAGCATTTAATCCTTTACATATTCATATGACAAAGACCTTGTTCACATGTTGTCT	720
Qy	721	GCTGACCCCTCCGCCACAATTTGCTCTTGACCCCTGACACATCCGCCCTTGAGAAACAC	780
Db	721	GCTGACCCCTCCGCCACAATTTGCTCTTGACCCCTGACACATCCGCCCTTGAGAAACAC	780
Qy	781	CCACGATGATCATGTAATACTAAGGGAACTCAGAGCGTGGCGGAGTCTCCATATGCTG	840
Db	781	CCACGATGATCATGTAATACTAAGGGAACTCAGAGCGTGGCGGAGTCTCCATATGCTG	840
Qy	841	AACGCTGGTCCCGGGGTCCTTCTTTCTTTCTTATAT	879
Db	841	AACGCTGGTCCCGGGGTCCTTCTTTCTTTCTTATAT	879
RESULT 4			
AF074086			
LOCUS	AF074086	19221 bp	DNA linear PRI 11-MAR-2002
DEFINITION	Homo sapiens tandemly repeated human endogenous retrovirus HERV-K (HML-2, HOM), complete sequence.		
ACCESSION	AF074086		
VERSION	AF074086.2	GI:9558700	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 19221) Mayer,J., Sauter,M., Racz,A., Scherer,D., Mueller-Lantzech,N. and Meese,E.		
TITLE	An almost-intact human endogenous retrovirus K on human chromosome 7		
JOURNAL	Nat. Genet. 21 (3), 257-258	(1999)	
MEDLINE	99178257		
PUBMED	10080172		
REFERENCE	2 (bases 1 to 19221) Reus,K., Mayer,J., Sauter,M., Scherer,D., Muller-Lantzech,N. and Meese,E.		
AUTHORS			
TITLE	Genomic organization of the human endogenous retrovirus HERV-K(HML-2, HOM) (ERV6) on chromosome 7		
JOURNAL	Genomics 72 (3), 314-320	(2001)	
MEDLINE	11401447		
PUBMED	21295053		
REFERENCE	3 (bases 1 to 19221) Mayer,J., Mueller-Lantzech,N. and Meese,E.U.		
AUTHORS			
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUN-1998)	Institut fuer Humangenetik, Universitaet des Saarlandes, Homburg 66421, Germany	
REFERENCE	4 (bases 1 to 19221) Reus,K., Mayer,J. and Meese,E.U.		
AUTHORS			
TITLE	Direct Submission		
JOURNAL	Submitted (25-JUL-2000)	Institut fuer Humangenetik, Universitaet des Saarlandes, Homburg 66421, Germany	

REMARK
Sequence update by submitter
COMMENT On Jul 28, 2000 this sequence version replaced gi:4456987.
FEATURES
Source
1. 19221
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/clone_id="Y20-B19"
/clone_lib="Lawrence Livermore National Laboratory Single
Chromosome Library: L107NC01"
1. 17976
/note="HML-2, 'HOM', tandemly repeated"
/rpt_family="Human endogenous retrovirus HERV-K"
/rpt_type=dispersed
1. 968
971. .988
/note="lysine primer binding site"
112. .3112
/gene="gag"
1112. .3112
/gene="gag"
/codon_start=1
/product="gag protein"
/protein_id="AF88165.1"
/db_xref="GI:9558702"
/translation="MGQTSKIKSKYASTYLSPIKILLRGGVSTRNLIKLPQIEQ
FCWPEPGGTLLDKWKRIGKELKQAKRKNIIPLTVMDWALIKALKEPTEDESV
SVSDAGSGCIIIDCNWTRKSKQKETEGLHCEVAEPMVASTQNVYNQIQEVIYPT
LKKEGSELPVAGSESKPRGTPSLPAGQVPTLQPKOVENKTOPVAQVPPAL
QYRPPESSGYGMPGMPAPQGRAPYPOPTPRRLNTPAPRSGSKLHEIIDSKRED
TEAWQPVLTLEPMRPGEGAGSESPVTEARYSFSIKLKMKEGYKQVQPNPYMT
LDSILAHGRHLIPIYDEILAKSSLSPSQFLQFTTWIIDGVEQVRRRRANPNYND
DQLGLGQNWSTISQQLMONEAIEQVRAICLAWERIDQPSGTCSEFNTVRGSSKEP
YVPCFVARLDQVAQKSIADKARKYIVELAWENANBECOSAIPLKGVKPAQDLYSE
YVACGKIIIGGAMKALMAQALRGVILGCGVRFPGKCYACGIGLKKKCPYLNKQ
ITIQATTTGHEPDLCPCKCKGHWASQCRSKDPKXGQPLSGNBQGPQAPOQTGAF
PIQEPYQGRGQQLPISQYFGQISQLPQYNCPPOAAVQ"
1. 3918
23915. 3918
/note="translated by frameshift"
/codon_start=3
/product="protease"
/protein_id="AF88166.1"
/db_xref="GI:9558703"
/translation="WATVGEKRAKPAQSGFTTWGVIIPNSAICSSGSGTPTTPVPSVS
GNRPVTTIQLSPATSGSAVDLCITQVSLIGEBEPQKPTVSGLPKGTGVLIG
RSLSLNKGVOIHRSVSDSYKGEILOVISIIPMSAPRDIRIQLLLPYIKGNSBI
KRIIGSGSDPTPKAAYMASOVSENRPVCAKIIQGFQEGVLTDGADVSIIALNMPK
MWKOKAVGVLGVTIGTASEVYSTETILHCLGPNOSTQVPMITSIPLINMGDILQQ
WGAETNPAPSYSPTSQKIMTKMGYIPGKGLGNBQGIKIPVAKINQEBGIGNPC"
1. 6749
3878. .6749
/gene="pol1"
3878. .6749
/gene="pol1"
/note="translated by frameshift"
/codon_start=2
/product="polymerase"
/protein_id="AF88167.1"
/db_xref="GI:9558704"
/translation="NKSRRKRNRBSLLGATVPEPKPIPLTWKTEKPVWYNQWPLPKO
KLEAHLHLANEQLEKHIEPSRPNAPSPTVIOKSGSKRMULDRAVNAVIOGSEL
QPELSPAMIPKOWPLIIDLKDCFTPIABEDQCEFAFATIPAINNKEPPTOMVY
LQPGMNSPTICQTPVGRALQVREKFSDEYIHICIDIDLCAETDKDLIDCTYFLQ
EVANAGLAISDKICISTEPHYGLMOIENKRIKPKIEIKDQIKQVQKQKLGJIN
WIRPTGIPYAMSNLSFSLRGSDSNLSKMLPEPAKELIKVEKLIQSOAQNRIPL
APLILIRTAHSPTGIIITQNDIIVMSPLPHSTVTFPTLYLQDIATLLIGCRRLRIK
LGNDNDPKIVPLITKQVROAFINSANMKIGLANPGLIDNHPKTKIPPLGLTWTI
LPKTRDEPLENALTVFDGSSNGKAYTGPKRVIKTPYQSAQRAELVAVITLQDF
DOJINIISSAVVQAQTRDVETALIRYSMDQDQNLQNLQOTVRKRNFPFYTHIRLA
HTNLPEPLTFKANEQADLLVSALIKAOELHALTHVNAAGLKNEDVTWKGADKIVQHC
TQOOVGLHTPOEAGVNRGLCPNALQMDVTPVTSFGRLSYVATVDTYSHFIATQIC
TGESSTHVHKHLISCPANVGPERKITDNGPVCSCAPKQFESQMKISHTTGYPYNDQ
GQAVIRINRITKTLQVKQKGGSDSECTTPQMLNALATLTLANIYRQQTTSAQ


```
HLTGKNSPHEGKLIIMKDNKNTWEIGKVIITWGRFACVSPENOLPVMIPTRHKE
YNEPIDAKKSTAEETSSQSTVDSQDEQNGVRRTDEVAIHQEGRAANLGYTKEAD
AVSYKISREHKGTNPREVAAACSLDCCINGKSPYACRSCSS"
6451..8550
/gene="env"
6451..8550
/gene="env"
/codon_start=1
/product="envelope"
/db_xref="GI:9558705"
/translation="MNPSEMQRKAAPRRRRHRNRAPIITHKNKNTVISEEQMLPSTYK
AEPPTWAKLKLQLATKYLENTVQTPESSMLAALMIVSMVLSLPMGAAANNT
YMAVPPPLIRAVTMDNPEYVNDVSWVPGPIDRCPAKPEEGEMINI SIGYH
PPIICGAPGLMAPVQNMVLVEVPTVSPICRPTTHMVSGLMRVRYLQDFYSRL
KPRKGRCPCKEIPKESKRTBVLMCEVANSANITLONNEFGTIDMAPRQPHNS
GOTQSCSAQVSPAVSDLTESLDKHKKLSQSTYPRMGEKGI STREPKIVSPSP
EHEBLMLTVASHHRIIMSGNQLTETDRKRFYTI DLNSSLTVPLQSVKPPYMLVVG
NIVAKPDSQITTCENCRLTICIDSTFNMQRHILVPRSGEWMI PVSMDRMEASPSVH
ILTEVLGVLNRSKRPIETLIVIMGLIATATAVAGVALHVSQSVSNVMDQKNS
TRLNMSQSIDOKLANOINDROTIVMGMRLMSLHRPOLCDDMTSDPCITPOIYN
ESHEHMDVRHROGRDNTLIDISKLEQIFEASKAHNLVBTETALVAGVADLANL
NPVTWKTIGSTITLILILVCLFCLLVCRCTQULRSDHKERAMMTAAVLSKSK
GGNVGSKRDQIVTVSV"
8505..9472
9475..9492
/feature="lysine primer binding site"
9578..9586
/feature="splice donor site"
9616..11616
/gene="gag"
9616..11616
/gene="gag"
/product="gag"
/protein_id="AAD21095.2"
/db_xref="GI:9558701"
/translation="MGQTSKIKSKKASVLSYFIKILKRGGVKVTNLIKLPQIIBO
FCPRFPGQTILDKMKRIGKELKQAKRKNIITLVTWMAIITKALPEPCEBDSV
SVSDAPGSCITIDCNENTRKSKQKTBGHCYVAFVMAOSTVDVNOLOEVIYFET
LKLEGPBELVGPBESKRGRTSPLPAQVPTLQPOKOVENKTQPPVAQOVAPAL
QYRPPESQYVGPMPAPQGRAPVPOPTRLNPTAPSPKQSKLHEIIDKSKED
TEAMQFVLTLEPMPGEAGQGEPTVEARYSIKLMDKMGVQYQNSPYMT
LLOSILANGHLIYDWEILIAKSSLSPOGLFQFTWMLIDVQOVRRAANPVNITA
DOLLIGONNSTISQALMONEALEQVRAICTRAWEKIDOBGSCSPFNVRQSKER
YVDAVARLDVAKSILADEKARKVIELEMAENANPBCQSIKPLKGVKSPVISE
YVKACDGIGAMKAMLAQAITGVVLGGVTRGRKCYNGQIGHLKQCPVLNKN
ITIGATTGTRPDLPCRCKGKMAQCSKSKPKNGQPLSGNFRQGPAPQQTGA
PIOPFVPOGFGQOPPLSQVYQGISQLPYVNNCPPPAAVQO"
<11419..12422
/feature="translated by frameshift"
/codon_start=3
/product="protease"
/protein_id="AAD21096.1"
/db_xref="GI:4456989"
/translation="WATVGRKAKPASPPTWNGIPNSATCSGSGTTPYPSVS
GNKPVTTIQLSPATSGSAVDLCTIQVLSLPGPPQKPTTGYGDLPKRTVLIIG
RSSLNLKGQVQHTSVDSVDYKGEIQVLISSIPMSASPRDRIADLLLPYKGNSEI
KRIQGLGSDPTGKAAVWASQVSENRPVCKAIIQKQFEGVLDGAVSIIALNMPK
NMPKQAVTGLVIGTASVYQSTEILHCLGPDNCGVMTSISPLNLMGRDLQO
WGAEITMPAPSPTSQIMTKMGYIPGKLGKNEGDKIPVEAKINQBERGICNP"
<12382..15253
/gene="pol"
<12382..15253
/gene="pol"
/feature="translated by frameshift"
/codon_start=2
/product="polymerase"
/protein_id="AAD21097.1"
/db_xref="GI:4456990"
/translation="NKSRRNRRESLLGAATVEBPPIPLTWKTEKPVWQNPPLQK
KLKLAHLIANQLKRGHIESPSPMSNVFVFIQKSGKGRMLTDLRAVNAVIOGPI
QGLPSRPMIRKMDPLIITIDKCFITPLAEOGCKKAFITIPAINNKEPATRQWKY
```

```
LPQMLNSPTTCQTFVGRALQVREKPSDCYIIHCIDILICAAETKDKLIDCTFLOA
EVANAGLIAIASDKIQSTSPHYKQMOIENKIKRQKEIKRQDLITNDQKILGIDN
WIRPTQIPYANSLNFIIRGSDSDNSKMEIKELVEEKIQSOIKRIDLPL
APOLLIIPYAPHSPTGIIIONTLVEMSPFLSHFTVFTFLDIQALILQOTRLIK
LQCNDDIPATVHLTXEVOQAFINSQAMKIDFNHFKTIFQFLKTLTWI
LPKITTREPLENALTPTDSSNGKAYTPKERVITKTPYQSOABETVAVITLQDF
DQIPINIIISDAVVOQATRDVEYALIRYSMDQNLQFNLQOTVRKRNPFYITHTA
HTNLPGPLTKANEQADILVSSALIKAOEHALTHVNAAGKRNKEDVTWQAKDIQHC
TQCOVHLPTQOAGVNPBGICPNALMQMDLTVHPSFGRLSYVHTVYSHFTWTCO
TGESHSVKKHLISCFAVNGVPEKIKTDNPGVCSAFOKLSOMKISHTTGIPYNSQ
GOAIVRTNRNTLTQOLVKQEGGDSKECTPMQMLALYTLNFINVYRQTTSAEQ
HLTGKNSPHEGKLIIMKDNKNTWEIGKVIITWGRFACVSPENOLPVMIPTRHKE
YNEPIDAKKSTAEETSSQSTVDSQDEQNGVRRTDEVAIHQEGRAANLGYTKEAD
AVSYKISREHKGTNPREVAAACSLDCCINGKSPYACRSCSS"
14955..17054
/gene="env"
14955..17054
/gene="env"
Query Match 100.0%; Score 879; DB 9; Length 19221;
Best Local Similarity 100.0%; Pred. No. 1.1e-262;
Matches 879; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CDS
gene
61 AGGAGATCCATTTTGTATGACTAGAAAANTCTTCGCGCTTGAGTTCTGTATC 120
61 AGGAGATCCATTTTGTATGACTAGAAAANTCTTCGCGCTTGAGTTCTGTATC 120
121 TATGACCTTACCCCAACCCCGTCTCTGAAAACATGCTGTCCTCACTCAGGTTAA 180
121 TATGACCTTACCCCAACCCCGTCTCTGAAAACATGCTGTCCTCACTCAGGTTAA 180
121 TATGACCTTACCCCAACCCCGTCTCTGAAAACATGCTGTCCTCACTCAGGTTAA 180
181 ATGATTAAGGCGGTCAGAGATGCTTTGTTAAACAGATCTTGAAGCAGATGCTC 240
181 ATGATTAAGGCGGTCAGAGATGCTTTGTTAAACAGATCTTGAAGCAGATGCTC 240
241 CTTAAGGTCATACCACTCCCTTAATCTCAAGTCCCAAGGACCAAAAACCTGGGAAAG 300
241 CTTAAGGTCATACCACTCCCTTAATCTCAAGTCCCAAGGACCAAAAACCTGGGAAAG 300
301 CCGGAGGACCTTCGCTCCTAGGAAAGCAGGATATGTCACAGTTCTCCCATGTGATAG 360
301 CCGGAGGACCTTCGCTCCTAGGAAAGCAGGATATGTCACAGTTCTCCCATGTGATAG 360
361 CTTGAAATATGCGCTCGTGGGAAAGGAAAGACCTGACCGTCCCGACCCGACACCCGTA 420
361 CTTGAAATATGCGCTCGTGGGAAAGGAAAGACCTGACCGTCCCGACCCGACACCCGTA 420
421 AAGGCTCTGCTGCTAGAGAGATTAGTAAAGAGAAAGAAATGCTCTTTCAGTTGAGACA 480
421 AAGGCTCTGCTGCTAGAGAGATTAGTAAAGAGAAAGAAATGCTCTTTCAGTTGAGACA 480
481 AAGGAGGAGCATCTGCTCCGCTGCTCCCGGAGCAATGGAATGTCTCGGTATAAACCC 540
481 AAGGAGGAGCATCTGCTCCGCTGCTCCCGGAGCAATGGAATGTCTCGGTATAAACCC 540
541 GATTGTATGCTCATCTTACTAGATAGAGGAAACCGCTTGAAGGCTGAGAGTGGACCT 600
541 GATTGTATGCTCATCTTACTAGATAGAGGAAACCGCTTGAAGGCTGAGAGTGGACCT 600
601 GCGGCGAGCAATATCTGCTTTGTAAGCACTGAGATGTTTATGTATGATATCTTAAAG 660
601 GCGGCGAGCAATATCTGCTTTGTAAGCACTGAGATGTTTATGTATGATATCTTAAAG 660
661 CACAGCACTTAATCTTAACTGCTATGATGGAAGACCTTGTGTACATGTTGTCT 720
661 CACAGCACTTAATCTTAACTGCTATGATGGAAGACCTTGTGTACATGTTGTCT 720
721 GCTGACCTCTCCCAACATGTCTGTGACCTGACATCCCTCTTGAAGAAACAC 780
```



```
repeat_region /rpc_family="MALR"
8148..8458
/rpc_family="Alu"
repeat_region 8461..8747
/rpc_family="Alu"
repeat_region 8755..8806
/rpc_family="GGGA)n"
repeat_region 9349..9499
/rpc_family="L1"
repeat_region 9513..9762
/rpc_family="Alu"
repeat_region 9768..9936
/rpc_family="Alu"
repeat_region 9938..10243
/rpc_family="Alu"
repeat_region 10278..10576
/rpc_family="L1"
repeat_region 10577..10897
/rpc_family="Alu"
repeat_region 10898..11020
/rpc_family="L1"
repeat_region 11187..11411
/rpc_family="L1"
repeat_region 11412..11711
/rpc_family="Alu"
repeat_region 11712..11961
/rpc_family="L1"
repeat_region 11973..12264
/rpc_family="Alu"
repeat_region 12265..12610
/rpc_family="L1"
repeat_region 12611..12907
/rpc_family="Alu"
repeat_region 12908..13739
/rpc_family="L1"
repeat_region 13743..13876
/rpc_family="L1"
repeat_region 13885..13934
/rpc_family="MER1_type"
repeat_region 14077..14139
/rpc_family="MIR"
repeat_region 14576..14863
/rpc_family="Alu"
repeat_region 15075..15379
/rpc_family="Alu"
repeat_region 15648..16009
/rpc_family="Alu"
repeat_region 16015..16322
/rpc_family="Alu"
repeat_region 16326..16454
/rpc_family="Alu"
repeat_region 17256..17384
/rpc_family="L1"
repeat_region 17389..17628
/rpc_family="Alu"
repeat_region 17635..18045
/rpc_family="ERV.L"
repeat_region 18046..18454
/rpc_family="(TA)n"
repeat_region 18455..18515
/rpc_family="ERV.L"
repeat_region 18516..18576
/rpc_family="Alu"
repeat_region 18643..18952
/rpc_family="Alu"
repeat_region 19052..19447
/rpc_family="MALR"
```

```
Query Match 99.8%; Score 877.4; DB 9; Length 154037;
Best Local Similarity 99.9%; Pred.No. 4.1e-262;
Matches 878; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

1 TGTGGGAAAAAGCAAGAGATCAGATTGTTACTGTGTCTGTATGAAAGAGATGACAT 60

```
|||||
Db 47417 TGTGGGAAAAAGCAAGAGATCAGATTGTTACTGTGTCTGTATGAAAGAGATGACAT 47358
Qy 61 AGAGACTCCATTTTGTATGTAAGAAAAATTTCTGCGCTTGATTCGTATATC 120
Db 47357 AGGAGACTCCATTTTGTATGTAAGAAAAATTTCTGCGCTTGATTCGTATATC 47298
Qy 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTGTCCACTCAGGTTAA 180
Db 47297 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTGTCCACTCAGGTTAA 47238
Qy 181 ATGATTAAGGCGGTGACAGATGTCTTTGTAAACAGATGCTTGAAGCAGATGCTC 240
Db 47237 ATGATTAAGGCGGTGACAGATGTCTTTGTAAACAGATGCTTGAAGCAGATGCTC 47178
Qy 241 CTTAAGATCATACACATCCCTTAATCTCAAGTCCCAAGGACACAAAACCTGGGAAG 300
Db 47177 CTTAAGATCATACACATCCCTTAATCTCAAGTCCCAAGGACACAAAACCTGGGAAG 47118
Qy 301 CCGCAGGACCTCTGCTAGGAAAGCAGGTAATTGTCACAGTTCTCCCATGTGATAG 360
Db 47117 CCGCAGGACCTCTGCTAGGAAAGCAGGTAATTGTCACAGTTCTCCCATGTGATAG 47058
Qy 361 CTTGAATATGCTCTGCTAGGAAAGCAGGTAATTGTCACAGTTCTCCCATGTGATAG 420
Db 47057 CTTGAATATGCTCTGCTAGGAAAGCAGGTAATTGTCACAGTTCTCCCATGTGATAG 46998
Qy 421 AAGGCTCTGCTAGGAGATTTGTAAAGAGAGAGATGCTCTTTCAGTTGAGACA 480
Db 46997 AAGGCTCTGCTAGGAGATTTGTAAAGAGAGAGATGCTCTTTCAGTTGAGACA 46938
Qy 481 AGAGAAAGGCAATGTCCTGCTGCGGCAATGTCCTGCGATTAAGAACCC 540
Db 46937 AGAGAAAGGCAATGTCCTGCTGCGGCAATGTCCTGCGATTAAGAACCC 46878
Qy 541 GATTGTATGCTCATCTACTAGATAGAGAAAAACCGCTTAGGCTGAGGTGGAGCT 600
Db 46877 GATTGTATGCTCATCTACTAGATAGAGAAAAACCGCTTAGGCTGAGGTGGAGCT 46818
Qy 601 GCGGCGACCAATACTGCTTTGTAAAGCACTGAGATGTTATGTATGATATCTAAAG 660
Db 46817 GCGGCGACCAATACTGCTTTGTAAAGCACTGAGATGTTATGTATGATATCTAAAG 46758
Qy 661 CACAGCACTTAATCCTTTACATGTCATATGATGAGAAACCTTGTCACATGTTGCT 720
Db 46757 CACAGCACTTAATCCTTTACATGTCATATGATGAGAAACCTTGTCACATGTTGCT 46698
Qy 721 GCTGACCTCTCCCAACATTTGTTGACCCCTGACACATCCCTCTTGAGAAACAC 780
Db 46697 GCTGACCTCTCCCAACATTTGTTGACCCCTGACACATCCCTCTTGAGAAACAC 46638
Qy 781 CCACAGATGATCAATAAATCTAAGGAACTCAGAGCTGGCGGAGATCTTCATATGCTG 840
Db 46637 CCACAGATGATCAATAAATCTAAGGAACTCAGAGCTGGCGGAGATCTTCATATGCTG 46578
Qy 841 AAGCTGTCTCCCGGATCCCTCTTCTTCTCTATA 879
Db 46577 AAGCTGTCTCCCGGATCCCTCTTCTTCTCTATA 46539
```

RESULT 6
LOCUS HEN17834 8300 bp DNA linear VRL 21-OCT-1999
DEFINITION Human endogenous retrovirus K (HERV-K) elements, clone SD-C7-34LTR.
ACCESSION Y17834
VERSION Y17834.1 GI:4185945
KEYWORDS env gene; env protein; gag gene; gag protein; long terminal repeat;
SOURCE Human endogenous retrovirus K
ORGANISM Human endogenous retrovirus K
REFERENCE 1
AUTHORS Tonjes, R. R., Czauderna, F. and Kurth, R.

QY 721 GCTGACCTCTCCCAATTTGTTGACCTGTACACATCCCTCTTCAAGAAAC 780
DB 8145 GCTACCTCTCCCAATTTGTTGACCTGTACACATCCCTCTTCAAGAAAC 8204
QY 781 CCACAGATGATCAATTAATTAAGGAAGTCAAGAGCTGGCGGATCTTCATATGCTG 840
DB 8205 CCACAGATGATCAATTAATTAAGGAAGTCAAGAGCTGGCGGATCTTCATATGCTG 8264
QY 841 AACGCTGTTCCCGGCTCCCTCTTCTTCTCT 876
DB 8265 AACGCTGTTCCCGGCTCCCTCTTCTTCTCT 8300
RESULT 7
AF490464
LOCUS 8708 bp DNA linear PRI 08-APR-2002
DEFINITION Homo sapiens HERV-K' long terminal repeat, complete sequence; and
AF490464
envelope glycoprotein gene, partial cds.
AF490464.1 GI:20067082
ACCESSION
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 8708)
Reus, K., Stuhler, T., Mayer, J. and Meese, E. U.
Haplotype analysis of Human Endogenous Retroviruses: A genetic
variant of HERV-K'(HML-2.HOM) with an intact YXDD motif of reverse
transcriptase
Unpublished
2 (bases 1 to 8708)
Reus, K., Stuhler, T., Mayer, J. and Meese, E. U.
Direct Submission
Submitted (06-MAR-2002) Humangenetik, Universitaet des Saarlandes,
Kirtbergstrasse, Homburg 66421, Germany
FEATURES
source
1..8708
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/note="from Asian sample
endogenous virus: HERV-K'."
239..+8708
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/note="endogenous_virus: HERV-K'; from Asian sample"
239..1206
1350..1350
/note="nonfunctional gag protein due to mutation"
3153..4155
/note="nonfunctional protease protein due to mutation"
4115..6986
/note="nonfunctional polymerase protein due to mutation in
gag"
6688..+8708
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AA09535.1"
/db_xref="GI:20067083"
/translation="MNPSEMQKAPRRRRHRNAPLTHKNNKVTSEOMKLPSTK
AEPYMAQLKLTQLATKYLENTVOTPESMLAALMIVSVSLPMPAGAAANT
YMAVPEPPLRAVMDNDPEVNVDSVVPIDRCAPKEEEMINISGYH
PPIGIRAPGLMAYONMLVEVPTVSPICRFTYHWSGMSIRPVVYLDDFSYRSL
KRPKGRPKKEIPKESKNTBVLMEEVANSANILONNEPRTIIDAPRGQPTNCS
GQTOCSQAVSPAVSDLTESLDKHKKQSFYPMWESKGISTRPKIVSPVSGP
GHEPLWLTVAASHIRIWSGNQLTETDRKPEYTIIDNSSLIVPLQCVKRPYMLVVG
NIVIKPSQITTCNCRLLTICIDSTFMQHRILIVARBEGVLPVSDRPEWASPSVH
ILTEVLKGVNRSKRFIFTLIAVIMGLIAVTATAVAGVALHSVGSVNFVNDMPKNS

ORIGIN
Query/Match 99.6%; Score 875.8; DB 9; Length 8708;
Best Local Similarity 99.8%; Pred. No. 1e-261;
Matches 877; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
TRUNSSSIDOKLANOINDLRQTVIMWGBRLMSIEHRFOLQCDMDTSDPCITPQIYN
ESEHMDMVRHRLOQREDNLTLDISKSEKHLINVPETEAIAVADGIANL
NPTWVXTISTTIIINILIVLCFLLIVCRCTQQRSDHRRAMT"
QY 1 TGTGGGAAAAAGCAAGAGATCGATTGTTACTGTGTCTGTGTAGAAAAGTAGACAT 60
DB 239 TGTGGGAAAAAGCAAGAGATCGATTGTTACTGTGTGTGTAGAAAAGTAGACAT 298
QY 61 AGAGACTCCATTTTGTATGTAAGAAAATTTCTTGCCCTTGAGATTCTGTATC 120
DB 299 AGAGACTCCATTTTGTATGTAAGAAAATTTCTTGCCCTTGAGATTCTGTATC 358
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAAATGTCGTGTGTCACACTCAGGATTAA 180
DB 359 TATGACCTTACCCCAACCCCGTCTCTGAAAATGTCGTGTGTCACACTCAGGATTAA 418
QY 181 ATGATTAAAGGCGGTGAGAGATGCTTTGTTAAAGATGCTTGAAGGACATGCTC 240
DB 419 ATGATTAAAGGCGGTGAGAGATGCTTTGTTAAAGATGCTTGAAGGACATGCTC 478
QY 241 CTTAAGAGTATCAACACCTCCATATCTCAAGTACCGAGGAGACAAAACCTGGGAAG 300
DB 479 CTTAAGAGTATCAACACCTCCATATCTCAAGTACCGAGGAGACAAAACCTGGGAAG 538
QY 301 CCGCAGAGGACCTCTGCTCTAGAAAAGCAGATATTGTCACAGTTCTCCCATGTGATAG 360
DB 539 CCGCAGAGGACCTCTGCTCTAGAAAAGCAGATATTGTCACAGTTCTCCCATGTGATAG 598
QY 361 CTTGAATATGCTCTCTGCTGAGAGAGAAAGACTGACCGTCCCGACCGGACACCGGTA 420
DB 599 CTTGAATATGCTCTCTGCTGAGAGAGAAAGACTGACCGTCCCGACCGGACACCGGTA 658
QY 421 AAGGCTGTGCTGAGAGAGATTAAGTAAAGAGAAAGAAATGCTCTTGACGTTGAGACA 480
DB 659 AAGGCTGTGCTGAGAGAGATTAAGTAAAGAGAAAGAAATGCTCTTGACGTTGAGACA 718
QY 481 AGAAGAGAGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 719 AGAAGAGAGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 778
QY 541 GATTGTATGCTCAATCTAATGATAGAGAGAAACCCCTTAGGGCTGAGAGTGGACCT 600
DB 779 GATTGTATGCTCAATCTAATGATAGAGAGAAACCCCTTAGGGCTGAGAGTGGACCT 838
QY 601 GCGGGCAGCAATCTGCTTTGTAAAGACAGAGATGTTATGCTATAGCATATTAAG 660
DB 839 GCGGGCAGCAATCTGCTTTGTAAAGACAGAGATGTTATGCTATAGCATATTAAG 898
QY 661 CACAGACTTAATCTTATCATTTGATGATGCAAGACCTTGTTCACATGTTGTCT 720
DB 899 CACAGACTTAATCTTATCATTTGATGATGCAAGACCTTGTTCACATGTTGTCT 958
QY 721 GCTGACCTCTCCCAATTTGTTGACCTGTACACATCCCTCTTCAAGAAAC 780
DB 959 GCTGACCTCTCCCAATTTGTTGACCTGTACACATCCCTCTTCAAGAAAC 1018
QY 781 CCACAGATGATCAATTAATTAAGGAAGTCAAGAGCTGGCGGATCTTCATATGCTG 840
DB 1019 CCACAGATGATCAATTAATTAAGGAAGTCAAGAGCTGGCGGATCTTCATATGCTG 1078
QY 841 AACGCTGTTCCCGGCTCCCTCTTCTTCTCTATA 879
DB 1079 AACGCTGTTCCCGGCTCCCTCTTCTTCTCTATA 1117
RESULT 8
HEN17832
LOCUS HEN17832 10029 bp DNA linear VRL 21-OCT-1999

DEFINITION Human endogenous retrovirus K (HERV-K) elements, clone C7.
ACCESSION Y17832
VERSION Y17832.2 GI:4581240
KEYWORDS env gene; env protein; gag gene; gag protein; long terminal repeat;
SOURCE Human endogenous retrovirus K
ORGANISM Human endogenous retrovirus K
VIRUSES; Retrovirdae; Retroviridae.
REFERENCE
AUTHORS Tonjes, R.R., Czauderna, F. and Kurth, R.
TITLE Expression of full-length human endogenous retrovirus type K
JOURNAL U. Virol. 73 (11), 9187-9195 (1999)
MEDLINE 99445825
PUBMED 10516026
REFERENCE
AUTHORS 2 Toenjes, R.R.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-1998) R.R. Toenjes, Paul-Ehrlich Inst.,
Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-63207 Langen, FRG
REMARK 3 (bases 1 to 10029)
REFERENCE
AUTHORS Toenjes, R.R.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1999) R.R. Toenjes, Paul-Ehrlich Inst.,
Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-63207 Langen, FRG
COMMENT On Apr 12, 1999 this sequence version replaced gi:4185937.
Related sequence Y10390.
FEATURES
source
location/Qualifiers
1..10029
/organism="Human endogenous retrovirus K"
/proviral
/mol_type="genomic DNA"
/specific_host="human"
/db_xref="taxon:45617"
/clone="C7"
/clone_1lb="Pl (Genome Systems, St. Louis, MO, USA)"
/note="Assigned to human chromosome 7"
1..83
/rpt_family="Alu"
243..426
/rpt_family="Alu"
427..1394
/note="5' LTR"
1397..1414
1538..3538
/gene="gag"
1538..3538
/gene="gag"
/codon_start=1
/product="gag protein"
/protein_id="CAA76878.1"
/db_xref="GI:4185938"
/db_xref="GOA:Q9YNA9"
/db_xref="SPTREMBL:Q9YNA9"
/translation="MGQTKSKIKSKVSYLSFIKILKRGVNVSTKNLTKLFOIIEQ
FCWPFPEQGTLDLDMKRIKELKQAGKKNIIPLTYNDMAIIKALBEFOEBSY
SVADPGSCIIDNENRKKSOKEBEGHEVYAPVMAOSTONVDVNOLOAYIPET
LKI.BKGKPELVGSPSEKSGRTSPLPAGOVVTLQPOKOVENKTPOPAVADYPAEL
QYRPPEQYGRGMPAPRAGRAPYRPPPTRRNLPTAPRQSKSHEITDKSKED
TEAMQFPVTLPEMPGEGAOGEPPVEAKYSFSIKLDMKEGVAVQVPSNYMT
LDSIAHGRLIPYDWEILAKSLSPQIQTWIDVQOVRRRRANRPVNI
DADLIGCONNSTISQALMONEAIEYRAICLAMEKIQPGSTCSFNTVRGSGSEP
YPPVAFALQVAKSIADEKARKVIELMAYENANPECGSAIPLKGVAGSDVISE
YVACDGIAGAMKAMMAQITGVITGGVTRPGRCVNCGOIGHKXKCPILANON
ITTOATTGTEPEPDLCPCKKAGKMAOCCKSKDKNQPLSGNBQKQPAPOQTGAF
PIQPFVQGFQGGQPPUSVVFQGISQLPQYNNCPQPAAVQ"

/protein_id="CAA76879.1"
/db_xref="GI:4185939"
/db_xref="GOA:Q9YNA9"
/db_xref="SPTREMBL:Q9YNA9"
/translation="MLTDLRAVNAVIOGPGLOGLPSPAMIPKMDPLIITLIDKCFE
TPIARODEKFAFTIPAINKEBPATRFQMKLPOGLINSPTICQFVSGALAPVEK
FSDCYITHCIDLILCAETKDKLIDCYTFLQAEVANAAGLAIASDKIOTSTPRHYLQ
IENRKIRPKQETIRKDTLTLNDFOKLDINNIRPLIGIPYAMSLFSLNGSDL
NSKMLTPENTATKELKVBESKIOAQINRIDLPLQLIFATASPGIILIQNTDVE
WSFLPHSTVTEFTLYDOJATLIGQRLRIILKCGNDPKIVPLTKEQVQAIFNSG
AKKIGANFVGIIDNHYPKTKIPOFLKTLIPLKTRREPLNALTVFQSGSNGA
AYNGPERVITKTPYOSAORAEIYAVTVAQEPDOPINIIISDSYVVQAFDVEITALIK
YSMDQDNLQNLQOTVTRKRPPIYTHIRATNIPGLITRANBQDLVSSALIK
QELHALTVNAAGLKNKFDVTWKQADIVQCTQCVLHPTBAGNPRGLCPNALM
QMDVTVPSEGRSLVYHVVDYSHPIWATQCESTSHYKHLCSFAVGPBEKIK
TDNGPGYCSKAFQKFLSOMKISHTTGI PYNSQOALIVERTRLTKTLQVKGSGDSK
ECTTPOMQNLALTYLNLPLNIYNOTTSABOHLTGKNSPHEGKLIIMWDSKNKTE
IGKVTIWRGGRFACVSRGNQLPWTIPTRHKFNEPIRAKSTSAETESQSTDS
QDRQNDVVRITDVAIHQGRANLQTTKADAVSYKISREHKGDTNPREYACSLDD
CINGKSPYACRSCS"
6892..8976
/gene="env"
6892..8976
/gene="env"
/gene="env"
/codon_start=1
/product="env protein"
/protein_id="CAA76880.1"
/db_xref="GI:4185940"
/db_xref="GOA:Q9YNA9"
/db_xref="SPTREMBL:Q9YNA9"
/translation="MQRKAPRRRRRRNAPLTHKKNKNTVTSQOKLPSTKAEPT
MAQLKLTOLATKYLENTVYQTPESMLLAALMIVSMVSVLPMRPAAGAAANVYMAV
PEPPLIRAVTMNDPTEVYVNDVSWVPGIDRCAPKEBGMWIMISIDYHPLCL
GRAPGLMAYQNVMLVEPVPSPICRFTHYVSGMLRPVNYLOFSYSLKFPBK
GRCPKEIRPESKNTFVLWBECEVANSAYLVIONEGTIIIDMAPROGFVPSGEBEPL
COSAOYSPAVDSLTESI.LKHKKKIOSYPMWMBGKGIIDPAPKIVSPVSGEBEPL
WRITVASHHRTISGQTLFETDRKFPYITLINSSTLVPLOSCVKKPYMLVGNITIK
PSQTLTNCNRLTCTIDSTFWQHRILLVRAEGVWIPVSMRPMWASVSHLIEV
LKGVLNRSKRIFFTLIVIMGLAVATVAVAGVALHSVQSVNFVNDKQNSRTRICL
SOSSIDOKLANQINDRQTVIWMGDRLMSLEHFOLOCDNNTSDFCITPQIYNESEH
MDVVRHLQGRBNLTLDISKLEQIFBASKALNIYVGEATAGVADGIANLPVYM
VKTIGSTTINIILIVICFLILVCRQQLRDRBDRBRAMMTAVLSKRGANGV
KSKRDIVTVSV"
8931..9898
/note="3' LTR"
9899..9903
/rpt_family="Alu"
ORIGIN
Query Match 99.5%; Score 874.2; DB 14; Length 10029;
Best Local Similarity 99.7%; Pred. No. 3.3e-261;
Matches 876; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TGTGGGAAAAGCAAGAGATCATGTTGCTGCTGTGTATAGAAAAGATAGCAT 60
Db TGTGGGAAAAGCAAGAGATCATGTTGCTGCTGTGTATAGAAAAGATAGCAT 486
QY 427 TGTGGGAAAAGCAAGAGATCATGTTGCTGCTGTGTATAGAAAAGATAGCAT 486
Db TGTGGGAAAAGCAAGAGATCATGTTGCTGCTGTGTATAGAAAAGATAGCAT 486
QY 61 AGGAGATCATTTTGTATGTAAGAAAATTTCTTGCTGCTGAGATTCTGTATATC 120
Db AGGAGATCATTTTGTATGTAAGAAAATTTCTTGCTGCTGAGATTCTGTATATC 546
QY 487 AGGAGATCATTTTGTATGTAAGAAAATTTCTTGCTGCTGAGATTCTGTATATC 546
Db AGGAGATCATTTTGTATGTAAGAAAATTTCTTGCTGCTGAGATTCTGTATATC 546
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACAATGCTGTGCTCACTCAGAGTTAA 180
Db TATGACCTTACCCCAACCCCGTCTCTGAAACAATGCTGTGCTCACTCAGAGTTAA 606
QY 547 TATGACCTTACCCCAACCCCGTCTCTGAAACAATGCTGTGCTCACTCAGAGTTAA 606
Db TATGACCTTACCCCAACCCCGTCTCTGAAACAATGCTGTGCTCACTCAGAGTTAA 606
QY 181 ATGATTTAAGGGGCGTGCAGAGATGCTTGTTTAAACAGATGCTTGAAGCAGATGCTC 240
Db ATGATTTAAGGGGCGTGCAGAGATGCTTGTTTAAACAGATGCTTGAAGCAGATGCTC 666
QY 607 ATGATTTAAGGGGCGTGCAGAGATGCTTGTTTAAACAGATGCTTGAAGCAGATGCTC 666
Db ATGATTTAAGGGGCGTGCAGAGATGCTTGTTTAAACAGATGCTTGAAGCAGATGCTC 666
QY 241 CTTAAGAGTCATCACCACCTCCTTATCTCAAGTACCAGGAGACACAAAACCTCGGAAAG 300
Db CTTAAGAGTCATCACCACCTCCTTATCTCAAGTACCAGGAGACACAAAACCTCGGAAAG 726

QY 301 CCGCAGGAGCCTCTGCTAGAGAAAGCAGGATTTGTCACAGTTCTCCCATGTGATAG 360
DB 727 CCCAGAGGACCTTCTGCTAGAGAAAGCAGGATTTGTCACAGTTCTCCCATGTGATAG 786
QY 361 CCGGAAATATGCGCTCTGAGGAAAGAAAGCCTGACCGTCCCGACCCGACACCCGTA 420
DB 787 CCGGAAATATGCGCTCTGAGGAAAGAAAGCCTGACCGTCCCGACCCGACACCCGTA 846
QY 421 AAGGCTGTGCTGAGGAGGATTTAGTAAAGAGAGAAATGCTCTTGAGTTGAGACA 480
DB 847 AAGGCTGTGCTGAGGAGGATTTAGTAAAGAGAGAAATGCTCTTGAGTTGAGACA 906
QY 481 AGAGAGAGCATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 907 AGAGAGAGCATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 966
QY 541 GATTGTATGCTCCATCTACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 967 GATTGTATGCTCCATCTACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026
QY 601 GCGGGAGCAATCTGCTTTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 1027 GCGGGAGCAATCTGCTTTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086
QY 661 CACAGACATTAACTCTTACATTTGTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 1087 CACAGACATTAACTCTTACATTTGTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1146
QY 721 GCTGACCTCTCCGACAAATTTGTTGATGACCTGACACATCCCTCTTTCAGAGAAACAC 780
DB 1147 GCTGACCTCTCCGACAAATTTGTTGATGACCTGACACATCCCTCTTTCAGAGAAACAC 1206
QY 781 CCACAGATGATCACTAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 1207 CCACAGATGATCACTAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266
QY 841 AAGCGTGTTCCTCCGAGGCTCCCTCTTCTTCTCTCTATA 879
DB 1267 AAGCGTGTTCCTCCGAGGCTCCCTCTTCTTCTCTCTATA 1305

RESULT 9
AF261945 6206 bp DNA linear PRI 26-SEP-2000
LOCUS AF261945
DEFINITION Homo sapiens human endogenous retrovirus K108, partial sequence.
ACCESSION AF261945
VERSION AF261945.1 GI:10304107
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 6206)
Turner,G., Barbulescu,M., Seaman,M.I., Kidd,K.K. and Lenz,J.
A genetic variant of HERV-K108 in humans with a YXDD Motif in
reverse transcriptase
Unpublished
2 (bases 1 to 6206)
Turner,G., Barbulescu,M., Seaman,M.I., Kidd,K.K. and Lenz,J.
Direct Submission
Submitted (28-APR-2000) Molecular Genetics, Albert Einstein College
of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA
Location/Qualifiers
1..6206
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
Source
repeat_region
48..6206
/rpt_family="human endogenous retrovirus K108"
48..1015
LTR
ORIGIN
Query Match 99.1%; Score 871; DB 9; Length 6206;

Best Local Similarity 99.4%; Pred. No. 3.2e-260;
Matches 874; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 TGTGGGAAAAGAGAGATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 60
DB 48 TGTGGGAAAAGAGAGATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 107
QY 61 AGAGAGATCTCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
DB 108 AGAGAGATCTCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 167
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 180
DB 168 TATGACCTTACCCCAACCCCGTCTCTGAAACATGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 227
QY 181 ATGGAATTAAGGGGGGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 228 ATGGAATTAAGGGGGGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 287
QY 241 CTTAAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 288 CTTAAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 347
QY 301 CCGCAGGAGCCTCTGCTTACGAG 360
DB 348 CCGCAGGAGCCTCTGCTTACGAG 407
QY 361 CCGGAAATATGCGCTCTGAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 408 CCGGAAATATGCGCTCTGAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 467
QY 421 AAGGCTGTGCTGAGGAGAGATTTAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 468 AAGGCTGTGCTGAGGAGAGATTTAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 527
QY 481 AAGAGAGAGCATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 528 AAGAGAGAGCATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 587
QY 541 GATTGTATGCTCCATCTACTAG 600
DB 588 GATTGTATGCTCCATCTACTAG 647
QY 601 GCGGGAGCAATCTGCTTTGTAAG 660
DB 648 GCGGGAGCAATCTGCTTTGTAAG 707
QY 661 CACAGACATTAACTCTTACATTTGTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 708 CACAGACATTAACTCTTACATTTGTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 767
QY 721 GCTGACCTCTCCGACAAATTTGTTGATGACCTGACACATCCCTCTTTCAGAGAAACAC 780
DB 768 GCTGACCTCTCCGACAAATTTGTTGATGACCTGACACATCCCTCTTTCAGAGAAACAC 827
QY 781 CCACAGATGATCACTAAATTAAG 840
DB 828 CCACAGATGATCACTAAATTAAG 887
QY 841 AAGCGTGTTCCTCCGAGGCTCCCTCTTCTCTCTATA 879
DB 888 AAGCGTGTTCCTCCGAGGCTCCCTCTTCTCTCTATA 926

RESULT 10
HERVORF 1505 bp mRNA linear VRL 06-APR-1995
LOCUS X82271
DEFINITION Human endogenous retrovirus mRNA for central open reading frame.
ACCESSION X82271
VERSION X82271.1 GI:757869
KEYWORDS open reading frame.
SOURCE Human endogenous retrovirus
ORGANISM Human endogenous retrovirus

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE

1
Lower, R., Tonjes, R.R., Korbmacher, C., Kurth, R. and Lower, J.
Identification of a Rev-related protein by analysis of spliced
transcripts of the human endogenous retroviruses HTDV/HERV-K
J. Virol. 69 (1), 141-149 (1995)
7983704
2 (bases 1 to 1505)
Tonjes, R.R.
Direct Submission
Submitted (20-OCT-1994) R.R. Tonjes, Paul-Ehrlich Inst.,
Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-63207 Langen, FRG
Location/Qualifiers
1. 1505
/organism="Human endogenous retrovirus"
/mol_type="mRNA"
/specific_host="Homo sapiens, patient GH, teratocarcinoma"
/db_xref="taxon:11827"
/note="HERV-K"
1. 1505
/gene="unknown orf"
255. 256
/gene="unknown orf"
/note="splice site"
273. 590
/gene="unknown orf"
/codon_start=1
/protein_id="CAA57722.1"
/db_xref="GI:757870"
/db_xref="SPTREMBL:Q69383"
/translation="MNPSEMQKAPRRRRHRRNAPLTHKKNKVTSEQMKLPSTK
AEPPTMAQLKLTQLATKYLENTKVTQTPESMLAALMIVMSAGVPSSEBATAIE
NGP"
436
/gene="unknown orf"
/note="additional nucleotide"
512
/gene="unknown orf"
/note="additional nucleotide"
533. 534
/gene="unknown orf"
/note="splice site"

ORIGIN

Query Match 99.0%; Score 870; DB 14; Length 1505;
Best Local Similarity 99.4%; Pred. No. 5.9e-260;
Matches 873; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTGTGTGTAGAAAGAGTACAT 60
Db 628 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTGTGTGTAGAAAGAGTACAT 687
QY 61 AGGAGATCCATTTTGTATGTACTAAGAAAATTCTTGCCTTGAGATTCTTATC 120
Db 688 AGGAGATCCATTTTGTATGTACTAAGAAAATTCTTGCCTTGAGATTCTTATC 747
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTCACCTCAGAGTTAA 180
Db 748 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTCACCTCAGAGTTAA 807
QY 181 ATGATTAAGGCGGTGAGAGTGTCTTGTAAAGATGCTTGAAGCAGATCTC 240
Db 808 ATGATTAAGGCGGTGAGAGTGTCTTGTAAAGATGCTTGAAGCAGATCTC 867
QY 241 CTTAAGAGTCATCAACCTCCCTTATCTCAAGTACCAAGGACCAAAAACCTGGGAAG 300
Db 868 CTTAAGAGTCATCAACCTCCCTTATCTCAAGTACCAAGGACCAAAAACCTGGGAAG 927
QY 301 CCGGAGGAGCTCTGCTAGGAAAAGCAGATTTGTCACAGTTTCTCCCATGTGAAG 360
Db 928 CCGGAGGAGCTCTGCTAGGAAAAGCAGATTTGTCACAGTTTCTCCCATGTGAAG 987

QY 361 CTTGAAATATGCGCTCTGTGGGAAAGGAAAGACTGACCGTCCCCAGCCGACCCGTA 420
Db 988 TCTGAAATATGCGCTCTGTGGGAAAGGAAAGACTGACCGTCCCCAGCCGACCCGTA 1047
QY 421 AAGGATCTGTGCTGAGAGAGATTAGTAAAGAGAAAGAAATGCTCTTGAGTTGAGACA 480
Db 1048 AAGGATCTGTGCTGAGAGAGATTAGTAAAGAGAAAGAAATGCTCTTGAGTTGAGACA 1107
QY 481 AGAGAAAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 1108 AGAGAAAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1167
QY 541 GATTGATGCTCATCTACTGATGAGATGAGAAACCGCTTAGGCTGAGTGGGACCT 600
Db 1168 GATTGATGCTCATCTACTGATGAGATGAGAAACCGCTTAGGCTGAGTGGGACCT 1227
QY 601 GCGGAGAGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 1228 GCGGAGAGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1287
QY 661 CACAGACTTAATCTTACATGCTATGATGATGATGATGATGATGATGATGATGATGAT 720
Db 1288 CACAGACTTAATCTTACATGCTATGATGATGATGATGATGATGATGATGATGATGAT 1347
QY 721 GCTGACCTCTCCCACAATTTGTTGTGACCTGACACATCCCTCTTGAGAAACAC 780
Db 1348 GCTGACCTCTCCCACAATTTGTTGTGACCTGACACATCCCTCTTGAGAAACAC 1407
QY 781 CCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
Db 1408 CCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1467
QY 841 AAGCGTGTCCCGGAGTCCCTCTTCTTCTCTAT 878
Db 1468 AAGCGTGTCCCGGAGTCCCTCTTCTTCTCTAT 1505

RESULT 11
HERVENV 3366 bp mRNA linear VRL 30-MAR-1995
LOCUS Human endogenous retrovirus env mRNA.
DEFINITION X82272.1 GI:757871
VERSION X82272.1 GI:757871
KEYWORDS env gene.
SOURCE Human endogenous retrovirus
ORGANISM Human endogenous retrovirus
VIRUSES; Retrovirus; Retroviridae.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE

1
Lower, R., Tonjes, R.R., Korbmacher, C., Kurth, R. and Lower, J.
Identification of a Rev-related protein by analysis of spliced
transcripts of the human endogenous retroviruses HTDV/HERV-K
J. Virol. 69 (1), 141-149 (1995)
7983704
2 (bases 1 to 3366)
Tonjes, R.R.
Direct Submission
Submitted (20-OCT-1994) R.R. Tonjes, Paul-Ehrlich Inst.,
Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-63207 Langen, FRG
Location/Qualifiers
1. 3366
/organism="Human endogenous retrovirus"
/mol_type="mRNA"
/specific_host="Homo sapiens, patient GH, teratocarcinoma"
/db_xref="taxon:11827"
/note="HERV-K"
1. 3366
/gene="env"
437. 2536
/gene="env"
/codon_start=1
/protein_id="CAA57723.1"
/db_xref="GI:757872"


```
/db_xref="GOA:Q69384"
/translation="MNPSEMRKAPRRPRRRHNRAPLTHKNNKVVTSBEQKLPSTKK
AEPPTAOLKULTOLATKYLENTKYTOTPESMLAALMTYSVTVSLMPRAAALANT
YMAVPPPLIRATVMDNPTVEVNVSWVPIDRCAPKEEEMINIISGYH
PPILGAPGLMAYONMLVEPTVSPICFTYHMYSGMLRPRVRYLIDFYSORL
KFRKGRPCPEKEIPKESKTEVLWEECVANASVILONNEGCTIDAPROCFYHNS
GOTOSCSAQSYPVSDLTESLDKHKHKKLQSFYPMWEGKISTPRKIVSPGSP
EHPLEMLTVASHHRIISMGNQLETDRKFPTVIDNLSLTVLOSQVPRVWVG
NIVKPSQRTICENCRLTCTDSTPNQRHILVVRBEGWITVSDRPMASPSH
ILTEVLGVLNRSKRF1FTLAVIMGLAVATAVAAGVALHSSVQSVNFNDQKNS
TRLWNSQSSIDQKLANQINDLRQTVIMWGDLSLEHRFOLQCDMNTSCITQIYL
ESEHMDMVRHLQGRBNLTLDISKLEQIFEASKAHMLNVPGETA1AGVADLANL
NPVTWKTIGSTTIINLILVCLFILLVCRCTQO1RRDSDHREARMTMTVA1SKXK
GNVYKSRKQD1VTVSV"
misc_feature
600
/gene="env"
/note="additional nucleotide"
676
/gene="env"
/note="additional nucleotide"
misc_feature
676
/note="additional nucleotide"
ORIGIN
Query Match 98.6%; Score 866.4; DB 14; Length 3366;
Best Local Similarity 99.3%; Pred. No. 8.4e-259;
Matches 870; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 TGTGGGGAAGCAAGAGATCAAGATTGCTGTCTGTGTAAGAAAGTAGACAT 60
DB 2491 TGTGGGGAAGCAAGAGATCAAGATTGCTGTCTGTGTAAGAAAGTAGACAT 2550
QY 61 AGGAGCTCCATTTGTATGTAAGAAATCTTCGCCCTTGAGATCTGTATC 120
DB 2551 AGGAGCTCCATTTGTATGTAAGAAATCTTCGCCCTTGAGATCTGTATC 2610
QY 121 TATGACCTTACCCCAACCCCGTCTCTGTGAACATGTGTCCTCACTCAGGCTTAA 180
DB 2611 TATGACCTTACCCCAACCCCGTCTCTGTGAACATGTGTCCTCACTCAGGCTTAA 2670
QY 181 ATGATTAAGGGCGGTGAGATGTCTTTTAAACAGATGCTTGAAGGACAGATGTC 240
DB 2671 ATGATTAAGGGCGGTGAGATGTCTTTTAAACAGATGCTTGAAGGACAGATGTC 2730
QY 241 CTTAAGAGTATCAACCTCCCTAATCTCAAGTACCCGAGACCAAAAATTGGGGAAG 300
DB 2731 CTTAAGAGTATCAACCTCCCTAATCTCAAGTACCCGAGACCAAAAATTGGGGAAG 2790
QY 301 CCGCAGGAGACTCTGCTTGAAGAAAGCCAGTATGTCCAAAGTTTCTCCCATGTATAG 360
DB 2791 CCGCAGGAGACTCTGCTTGAAGAAAGCCAGTATGTCCAAAGTTTCTCCCATGTATAG 2850
QY 361 CTTGAAATATGCTCTGTGGGGAAGGAAGACCTGACCGTCCCAAGCCGACACCCGTA 420
DB 2851 CTTGAAATATGCTCTGTGGGGAAGGAAGACCTGACCGTCCCAAGCCGACACCCGTA 2910
QY 421 AAGGCTCTGTGCTAGAGAGATTAGTAAAGAGAAAGAAATGCTCTTGGAGTTGAGACA 480
DB 2911 AAGGCTCTGTGCTAGAGAGATTAGTAAAGAGAAAGAAATGCTCTTGGAGTTGAGACA 2970
QY 481 AGAGGAAGGAGCATCTGTCTCTGCTGCTCCCTGGGCAATGAGATGCTCGATATAAACCC 540
DB 3091 AGAGGAAGGAGCATCTGTCTCTGCTGCTCCCTGGGCAATGAGATGCTCGATATAAACCC 3030
QY 541 GATTGTATGCTCATCTAAGATAGGAAAAACCCGCTTAGAGGCTGAGAGTGGAGCT 600
DB 3031 GATTGTATGCTCATCTAAGATAGGAAAAACCCGCTTAGAGGCTGAGAGTGGAGCT 3090
QY 601 GCGGGCAGCATATCTGCTTTGTAAGAGCATGATGTTTATGCTGATGATATCTAAAG 660
DB 3091 GCGGGCAGCATATCTGCTTTGTAAGAGCATGATGTTTATGCTGATGATATCTAAAG 3150
QY 661 CACAGCACTTAATCTTACATGCTTATGATGCAAGACCTTTGTTACATGTTTGTCT 720
```

```
DB 3151 CACAGCACTTAATCTTACATGCTTATGATGCAAGACCTTTGTTACATGTTTGTCT 3210
QY 721 GCTACCTCTCCCAACAAATGCTGTGAGACCCGAGACATACCCCTCTTGAGAAACAC 780
DB 3211 GCTACCTCTCCCAACAAATGCTGTGAGACCCGAGACATACCCCTCTTGAGAAACAC 3270
QY 781 CCACAGATGATCACTTAATCTAAGGAAGTCAAGAGCTGGCGGAGATCTCCATATGCTG 840
DB 3271 CCACAGATGATCACTTAATCTAAGGAAGTCAAGAGCTGGCGGAGATCTCCATATGCTG 3330
QY 841 AACCTGATGCTCCCGGGTCCCTTCTTCTTCTCT 876
DB 3331 AACCTGATGCTCCCGGGTCCCTTCTTCTTCTCT 3366
RESULT 12
AC025420 94145 bp DNA linear PRI 21-FEB-2003
LOCUS Homo sapiens 12 BAC Rpl1-491c17 (Rosewell Park Cancer Institute
DEFINITION Human BAC library) complete sequence.
ACCESSION AC025420
VERSION AC025420
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 94145)
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Primates; Catarrhini; Homnidae; Homo.
Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-oshman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbarta,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Doultwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guayara,W.,
Gunnarane,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O.,
Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homes,F.,
Howard,S., Huber,J., Huliy,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
Loulleged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Marondel,I., Martin,R.,
Martindale,A., Martinez,E., Massey,E., Mawliny,E., McLeod,M.P.,
Meador,M., Mei,G., Merscher,S., Metker,M., Miller,A., Miner,G.,
Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S.,
Ogih,M., Okuwon,G., Otagunye,N., Oyiedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojoubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
Shooshitari,N., Sisson,I., Sodergren,E., Sonake,T., Sparks,A.,
Stanley,H., Stone,H., Sutton,A., Svaltek,A., Tabor,P., Tamerisa,A.,
Tamerisa,K., Tang,H., Taney,J., Taylor,C., Taylor,T., Telford,B.,
Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
Washington,C., Watlington,S., Williams,G., Williamson,A.,
Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
Zorilla,S., Zuercherpacti,R. and Gibbs,R.
Unpublished
Direct Submission
TITLE
JOURNAL
REFERENCE
2 (bases 1 to 94145)
Worley,K.C.
```

TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Mar 1, 2001 this sequence version replaced gi:12656668.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----
Contig length: 94145
Phrap values in estimate: 93963
Average error rate (BCM-Phrap estimate): 4.17304e-05

Fraction of Phrap values less than 40 : 0.0184541
Number of consensus changing edits: 36
Number of N's in consensus : 0

Position	Original+Context	Edited+Context
5	ctctacaac (n) gnnccmnaa	ctctacaac (a) gacgtcgaa
7	ctnaacaang (n) ngcnmaaat	ctnaacaag (a) gcgtcgaaat
8	caacaangn (n) gcnmaaat	ctnaacaag (a) gcgtcgaaat
11	acaangn (n) nnaatcacg	acaagaagc (t) gcaatcacg
12	caangn (n) naatcacg	acaagaagc (g) caaatcacg
13	aangn (n) aatcacg	aaagaagc (c) aatcacg
33	ttctctta (n) ttctcttc	ttctctta (g) ttctcttc
36	ctctctta (n) ttctcttc	ctctctta (c) ttctcttc
47	ctctcttc (n) cagttcttc	ctctcttc (c) cagttcttc
2176	ctaatctc (n) cagttcttc	ctaatctc (c) cagttcttc
7320	tttggaga (n) agctctgc	tttggaga (g) agctctgc
7415	aattcttag (n) ggcacattt	aattcttag (t) ggcacattt
16116	ttagctcca (n) cctctgcct	ttagctcca (g) cctctgcct
16128	ctctgcctc (n) tgcctctcc	ctctgcctc (t) tgcctctcc
16139	tgctctctc (n) cagggmnn	tgctctctc (c) cagggmnn
16145	ctccncaagg (n) nncmnmty	ctccncaagg (t) tctccncaag
16146	ctccncaagg (n) nncmnmty	ctccncaagg (t) tctccncaag
16147	ccncaagg (n) cncmnmty	ccncaagg (t) tctccncaag
16149	ncagggmnn (n) ncmntggaa	ccagggctc (t) ccaatggaaa
16150	cagggmnn (n) cmtggaaa	ccagggctc (t) ccaatggaaa
16152	ggmnn (n) ntggaaaagt	gggtctctc (a) ctggaaaagt
43501	ggmnn (n) ntggaaaagt	gggtctctc (a) ctggaaaagt
43501	ggmnn (n) ntggaaaagt	gggtctctc (a) ctggaaaagt
48685	tggaataag (a) gaagtgtata	tggaataag (g) gaagtgtata
64342	actagaagg (n) gactgtctat	actagaagg (t) gactgtctat
64346	agtaattca (n) gttmctctt	agtaattca (t) gttmctctt
64347	atccangt (n) nctctnaa	atccangt (t) nctctnaa
64353	gtmctctt (n) aangmcca	gtmctctt (t) aangmcca
64356	nmctttaa (n) gcmccatag	nmctttaa (t) gcmccatag
64359	ttttaaagc (n) nccatagtg	ttttaaagc (c) nccatagtg
64360	tttaangn (n) ccatagttg	tttaangn (c) ccatagttg
64381	ctccataag (n) ccataccat	ctccataag (g) ccataccat
67952	gtagtaatta (t) ttcactatg	gtagtaatta (a) ttcactatg
71109	agagagaata (n) cncacttga	agagagaata (a) cncacttga
71111	agcagaatac (n) caactgcac	agcagaatac (a) caactgcac
88461	ttgtatatac (n) gttaatagag	ttgtatatac (a) gttaatagag

----- Distribution of Quality < 40 Bases -----

#	bases	5	10	15	20	25	30	35	40
1000	900	*							
800	800	*							
700	700	*							
600	600	*							
500	500	*							
400	400	*							
300	300	*							
200	200	*							
100	100	*							

Phrap Value Range

FEATURES

Location/Qualifiers
1. .94145
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"

Query Match 98.4%; Score 864.6; DB 9; Length 94145;

Best Local Similarity 99.0%; Pred. No. 4e-258;
Matches 870; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy 1 TGTGGGAAAAAGCAAGAGATGATGATTTGATCTGCTGTGTGTAAGAAAGATGACAT 60
Db 45648 TGTGGGAAAAAGCAAGAGATGATGATTTGATCTGCTGTGTGTAAGAAAGATGACAT
Qy 61 AGGAGACTCCATTTTGTATGATGATGATTTGATCTGCTGTGTGTAAGAAAGATGATGAT 120
Db 45708 AGGAGACTCCATTTTGTATGATGATGATTTGATCTGCTGTGTGTAAGAAAGATGATGAT 45767
Qy 121 TATGACCTTACCCCAACCCCGTCTCTGTAACATGCTGTGTGTAAGAAAGATGATGAT 180
Db 45768 TATGACCTTACCCCAACCCCGTCTCTGTAACATGCTGTGTGTAAGAAAGATGATGAT 45827
Qy 181 ATGATTTAAGGCGGTGAGATGCTTTGTGTAAGAAAGATGATGATGATGATGATGATGAT 240
Db 45828 ATGATTTAAGGCGGTGAGATGCTTTGTGTAAGAAAGATGATGATGATGATGATGATGAT 45887
Qy 241 CTTAAGAGTCATCAACCACTCCCTAATCTCAAGTACCCAGGAGCAAAAACTGGGGAAG 300
Db 45888 CTTAAGAGTCATCAACCACTCCCTAATCTCAAGTACCCAGGAGCAAAAACTGGGGAAG 45947
Qy 301 CCGGAGGAGACTCTGCTGAGAAAGCCAGATTTGCTCAACGTTTCTCCCATGTGATGATG 360
Db 45948 CCGGAGGAGACTCTGCTGAGAAAGCCAGATTTGCTCAACGTTTCTCCCATGTGATGATG 46007
Qy 361 CTTAATATATGCTCTGCTGAGAAAGAAAGCACTGACCTGCTCCCAAGCCGCAACCCGTA 420
Db 46008 CTTAATATATGCTCTGCTGAGAAAGAAAGCACTGACCTGCTCCCAAGCCGCAACCCGTA 46067
Qy 421 AAGGCTCTGCTGAGAGAGATTTAGTAAAGAGAGAGATGCTCTTTCAGTTGAGACA 480
Db 46068 AAGGCTCTGCTGAGAGAGATTTAGTAAAGAGAGAGATGCTCTTTCAGTTGAGACA 46127
Qy 481 AGAGAAAGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 46128 AGAGAAAGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 46187
Qy 541 GATTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 46188 GATTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 46247
Qy 601 GCGGCGAGCAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 46248 GCGGCGAGCAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 46307
Qy 661 CACAGCACTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCT 720
Db 46308 CACAGCACTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCT 46367
Qy 721 GCTGACCTCTCTCCCAACATTTGCTGTGACCTGACATCCCTCTCTTTCGAGAAACAC 780
Db 46368 GCTGACCTCTCTCCCAACATTTGCTGTGACCTGACATCCCTCTCTTTCGAGAAACAC 46427
Qy 781 CCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
Db 46428 CCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 46487
Qy 841 AACGCTGTTCCCGGCTCCCTCTCTTCTCTCTATA 879
Db 46488 AACGCTGTTCCCGGCTCCCTCTCTTCTCTATA 46526
```

RESULT 13
AC110927 181916 bp DNA linear HTG 20-FEB-2002
LOCUS Homo sapiens chromosome UNK clone Rpl11-599B11, WORKING DRAFT
DEFINITION
SEQUENCE, 14 unordered pieces.
AC110927
AC110927.2 GI:18767601
VERSION
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 181916)
Waterson, R.H.
The sequence of Homo sapiens clone

REFERENCE

2 (bases 1 to 181916)
Waterson, R.H.
Direct Submission

Submitted (16-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

3 (bases 1 to 181916)
Waterson, R.H.
Direct Submission

Submitted (20-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Feb 20, 2002 this sequence version replaced gi:18693568.

COMMENT

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: H_NH0599B11

Summary Statistics

Sequencing vector: Plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly: Dye-terminator Big Dye; version 0.990319
Consensus quality: 174138 bases at least Q40
Consensus quality: 176242 bases at least Q30
Consensus quality: 177435 bases at least Q20
Insert size: 18700; agarose-fp
Insert size: 180616; sum-of-contigs
Quality coverage: 0.00 in Q20 bases; sum-of-contigs
Quality coverage: 6.86 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1038: contig of 1038 bp in length
* 1039 1138: gap of unknown length
* 1139 2442: contig of 1304 bp in length
* 2443 2543: gap of unknown length
* 2543 4167: contig of 1625 bp in length
* 4168 4267: gap of unknown length
* 4268 5501: contig of 1234 bp in length
* 5502 5601: gap of unknown length
* 5602 9767: contig of 4166 bp in length
* 9768 9867: gap of unknown length
* 9868 16625: contig of 6758 bp in length
* 16626 16725: gap of unknown length
* 16726 21169: contig of 4444 bp in length
* 21170 21269: gap of unknown length
* 21270 27682: contig of 6413 bp in length
* 27683 27782: gap of unknown length
* 27783 38632: contig of 10850 bp in length
* 38633 38732: gap of unknown length
* 38733 54115: contig of 15383 bp in length
* 54116 66745: gap of unknown length
* 66746 66845: contig of 12530 bp in length
* 66845: gap of unknown length

AUTHORS
Waterston, R.
TITLE
Direct Submission
JOURNAL
Submitted (29-APR-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 205317)
AUTHORS
Wilson, R.
TITLE
Direct Submission
JOURNAL
Submitted (08-OCT-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 3, 2002 this sequence version replaced gi:15624987.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0518112
Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send
mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu>

SOURCE INFORMATION:
The RPCT-11 human BAC library was made from the blood of one male
donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frenken, E.,
Tateno, M., Catalano, J. J., and de Jong, P. J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genome 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.reagen.com>) or Pieter de Jong
and coworkers at <http://www.chori.org>
VECTOR: pBACe3.6.

FEATURES
Source
Location/Qualifiers
1..205317
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-518112"
/clone_1fb="RPCT-11"
3228..3365
/rpt_family="MER1_type"
3307..3391
/rpt_family="MER1_type"
3413..3520
/rpt_family="MER1_type"
3522..3552
/rpt_family="MER1_type"
3590..3727
/rpt_family="TGAG)n"
5024..5315
/rpt_family="(CCGGG)n"
/rpt_family="Alu"

repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region

5380..5433
/rpt_family="Alu"
5694..5878
/rpt_family="MIR"
6081..6519
/rpt_family="(TGG)n"
6558..6715
/rpt_family="(TGG)n"
6756..6853
/rpt_family="(TGG)n"
6875..7059
/rpt_family="(TGG)n"
7086..7949
/rpt_family="(TGG)n"
14663..14965
/rpt_family="(TGG)n"
16138..16164
/rpt_family="Alu"
17042..17107
/rpt_family="(A)n"
20446..20624
/rpt_family="(TGG)n"
20706..21006
/rpt_family="C-rich"
21038..21220
/rpt_family="C-rich"
21116..21847
/rpt_family="(CA)n"
/note="CpG island (4GC=62.6, o/e=1.44, #CpGs=49)"
21285..21466
/rpt_family="C-rich"
21454..21627
/rpt_family="C-rich"
21655..21805
/rpt_family="(CCCA)n"
21810..21830
/rpt_family="(CA)n"
21831..22358
/rpt_family="MER2_type"
22359..22657
/rpt_family="BRV1"
27420..27515
/rpt_family="MER2_type"
27526..27642
/rpt_family="(TGG)n"
27644..27823
/rpt_family="(TGG)n"
27834..27986
/rpt_family="(TGG)n"
28006..28078
/rpt_family="(TGG)n"
28080..28119
/rpt_family="T1"
28139..28275
/rpt_family="(TATATG)n"
30121..30429
/rpt_family="(TGG)n"
31275..31433
/rpt_family="Alu"
31447..31553
/rpt_family="MIR"
31578..31700
/rpt_family="L2"
33788..33688
/rpt_family="Alu"
34814..34986
/note="CpG island (4GC=66.0, o/e=0.70, #CpGs=62)"
35032..35199
/rpt_family="C-rich"
35215..35385
/rpt_family="C-rich"
35338..35510
/rpt_family="C-rich"
35434..35604
/rpt_family="CT-rich"
repeat_region

Query Match	98.4%;	Score 864.6;	DB 9;	Length 205317;
Best Local Similarity	99.0%;	Pred. No. 4,2e-258;		
Matches 870; Conservative	0;	Mismatches 9;	Indels 0;	Gaps 0;

QY	781	CCACGAGTATGATGAAATACCTAAGAGGAACTCAGAGCTGGCGGGAGTCCCTCATATCTG	840
Db	173394	CCACGAGTATGATGAAATACCTAAGAGGAACTCAGAGCTGGCGGGAGTCCCTCATATCTG	173453
QY	841	AAAGCTGGTCCCGGGGTCCCTTCTTTCTCTATA	879
Db	173454	AAAGCTGGTCCCGGGGTCCCTTCTTTCTCTATA	173492
RESULT 15			
LOCUS	AL139404		
DEFINITION	Human DNA sequence from clone RP11-376D11 on chromosome 10,		
ACCESSION	AL139404		
VERSION	AL139404.9		
KEYWORDS	GI:15617241		
SOURCE	HTG.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Bukhariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 122913)		
JOURNAL	Whitehead, S.		
COMMENT	Direct Submission Submitted (12-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiry: humquery@sanger.ac.uk request: clonerequest@sanger.ac.uk On Sep 13, 2001 this sequence version replaced gi:14586503. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr10 RP11-376D11 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBACe3.6		
FEATURES			
source	1..122913		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="10"		
	/clone="RP11-376D11"		
	/clone_1lb="RPCI-11.2"		
	/complement(18308..18730)		
	/note="match: GSS: EM:180175361"		
	complement(18360..18734)		
misc_feature			
misc_feature			

THIS PAGE BLANK (USP),

Web site: <http://www.shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdpax11.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
 Series: IRAX Plate: 48 Row: P Column: 5
 This clone has the following problem: retained intron.

FEATURES

source

1.1698
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5314057"
 /issue_type="Brain, hypothalamus"
 /clone_lib="NIH MGC_96"
 /lab_host="DH10B"
 /note="Vector: pBluescript"

ORIGIN

Query Match 94.0%; Score 826.2; DB 11; Length 1698;

Best Local Similarity 96.2%; Pred. No. 3e-241;
 Matches 846; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```

QY 1 TGTGGGAGAAAGCAAGAGATCATGATGTTACTGTGTCTGTGTAGAAAGATGACAT 60
DB 805 TGTGGGAGAAAGCAAGAGATCATGATGTTACTGTGTCTGTGTAGAAAGATGACAT 864
QY 61 AGGAGATCTCCATTTTGTATGATCTAAGAAAATTTCTTGCCTTGAGATCTTATATC 120
DB 865 AGGAGATCTCCATTTTGTATGATCTAAGAAAATTTCTTGCCTTGAGATCTTATATC 924
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGCTGTACCTCAGGGTTAA 180
DB 925 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGCTGTACCTCAGGGTTAA 984
QY 181 ATGATTAAAGGCGGTGAGATGTGTTTGTAAAGATGCTTGAAGGAGATGCTC 240
DB 985 ATGATTAAAGGCGGTGAGATGTGTTTGTAAAGATGCTTGAAGGAGATGCTC 1044
QY 241 CTTAAGAGTCATCACCACTCCCTTATCTCAAGTCCCAAGGACACAAAACCTGGGAAG 300
DB 1045 CTTAAGAGTCATCACCACTCCCTTATCTCAAGTCCCAAGGACACAAAACCTGGGAAG 1104
QY 301 CCGAGAGGACCTCTGCTAGGAAAGCAGGATGTCACAGTTTCTCCCATGTGATAG 360
DB 1105 CCGAGAGGACCTCTGCTAGGAAAGCAGGATGTCACAGTTTCTCCCATGTGATAG 1164
QY 361 CTTGAAATATGCTCTGTGGAGAAAGACCTGACCGTCCCGACCGGACACCCGTA 420
DB 1165 TCTGAAATATGCTCTGTGGAGAAAGACCTGACCGTCCCGACCGGACACCCGTA 1224
QY 421 AAGGCTCTGTGCTGAGAGATTTGTAAAGAGAGAAATGCTTCTTGACATTGAGACA 480
DB 1225 AAGGCTCTGTGCTGAGAGATTTGTAAAGAGAGAAATGCTTCTTGACATTGAGACA 1284
QY 481 AGAGGAAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 1285 AGAGGAAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1344
QY 541 GATTGTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
DB 1345 GATTGTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1404
QY 601 GCGGCGACGAATCTGCTTTGTAAAGCACTGAGATGTTATGTGTATGATATCTAAAG 660
DB 1405 GCGGCGACGAATCTGCTTTGTAAAGCACTGAGATGTTATGTGTATGATATCTAAAG 1464
QY 661 CACAGCACTTAATCTTATCTATGATGATGCAAGACCTTGTTCACATGTTTCTCT 720
DB 1465 CACAGCACTTAATCTTATCTATGATGATGCAAGACCTTGTTCACATGTTTCTCT 1524

```

QY 721 GCTGACCTCTCTCCCAACATTTGTGTGTGACCTGACACATCCCTCTTTCAGAAACAC 780
 DB 1525 GCTGACCTCTCTCCCAACATTTGTGTGTGACCTGACACATCCCTCTTTCAGAAACAC 1584
 QY 781 CCACAGATGATCGTAAATCTAAGGAACTCAGAGGCTGGGGGATCCCTCATATGCTG 840
 DB 1585 CCACAGATGATCGTAAATCTAAGGAACTCAGAGGCTGGGGGATCCCTCATATGCTG 1644
 QY 841 AACGCTGTTCCCGGGTCCCTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 879
 DB 1645 AACGCTGTTCCCGGGTCCCTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1683

RESULT 2

BU618928/c 706 bp mRNA linear EST 23-SEP-2002
 LOCUS BU618928
 DEFINITION UT-H-FH1-bf1-b-11-0-UI. s1 NCI CGAP FH1 Homo sapiens cDNA clone
 UT-H-FH1-bf1-b-11-0-UI 3', mRNA sequence.
 ACCESSION BU618928.1 GI:23285143
 VERSION BU618928
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 Tumor Gene Index

JOURNAL
 COMMENT Unpublished (1997)
 Contact: Robert Straubeberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: James Martin
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu
 The following repetitive elements were found in this cDNA sequence: 18-706, >LTR5#LTR/Retroviral (matched complement)
 Seq primer: M13 FORWARD
 PolyA+ies.

FEATURES

source

Location/Qualifiers
 1. 706
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-FH1-bf1-b-11-0-UI"
 /issue_type="Cell Line"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP FH1"
 /note="Organ: Chondrosarcoma; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site: 1: EcoR I; Site 2: Not I; NCI CGAP FH1 is a normalized cDNA library obtained from a cell line derived from grade I chondrosarcoma tissue. The library was constructed and normalized according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AGAATCCGCG. The cell line was provided by Dr. James Martin from the University of Iowa.
 TAG TISSUE=Human Chondrosarcoma Cell Line CS8 - Grade 1
 Chondrosarcoma
 TAG_LIB=UI-H-FH1
 TAG_SEQ=AGAATCCGCG"

ORIGIN

Query Match 75.0%; Score 659.2; DB 13; Length 706;
 Best Local Similarity 97.2%; Pred. No. 2.9e-190;
 Matches 670; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

```

QY 127 CTTACCCCAACCCCGTCTCTGAAACATGCTGTGTCTCACTCAGGGTTAATGAT 186
    |||
DB 706 CTTACCCCAACCCCGTCTCTGAAACATGCTGTGTCTCACTCAGGGTTAATGAT 647
    |||
QY 187 TAAGGGGGGTGACAGATGCTTTGTTAAACAGATGCTTGAAGCAGATGCTCTTAAG 246
    |||
DB 646 TAAGGGGGGTGACAGATGCTTTGTTAAACAGATGCTTGAAGCAGATGCTCTTAAG 587
    |||
QY 247 AGCATACACACATCCCAATCTCAAGTACCAGGACACAAAACCTGGAAAGGCGCAG 306
    |||
DB 586 AGCATACACACATCCCAATCTCAAGTACCAGGACACAAAACCTGGAAAGGCGCAG 527
    |||
QY 307 GGAAGCTCTGCTAGGAAAGCAGATGTTTCAAGCTTCTCCCATGTGATAGCTGAA 366
    |||
DB 526 GGAAGCTCTGCTAGGAAAGCAGATGTTTCAAGCTTCTCCCATGTGATAGCTGAA 467
    |||
QY 367 ATATGCGCTGTGGGAAAGGAAAGCCTGACCGTCCCGACCGGACACCCGTAAGGGT 426
    |||
DB 466 ATATGCGCTGTGGGAAAGGAAAGCCTGACCGTCCCGACCGGACACCCGTAAGGGT 407
    |||
QY 427 CTGTGCTGAGAGAGATTAATTAAGAGAAAGAAATGCTCTTGAAGTTGAGACAGAGA 486
    |||
DB 406 CTGTGCTGAGAGAGATTAATTAAGAGAAAGAAATGCTCTTGAAGTTGAGACAGAGA 347
    |||
QY 487 AGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 546
    |||
DB 346 AGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 287
    |||
QY 547 ATGCTCCATCTACTGAGATAGGAAAAACCGCTTGAAGGCTGAGAGTGGGACCTGCGGCG 606
    |||
DB 286 ATGCTCCATCTACTGAGATAGGAAAAACCGCTTGAAGGCTGAGAGTGGGACCTGCGGCG 227
    |||
QY 607 AGCAATCTGCTTTGTAAGCAGTGAATGTTATGTGTATGCTATCTAATTAACACAGC 666
    |||
DB 226 AGCAATCTGCTTTGTAAGCAGTGAATGTTATGTGTATGCTATCTAATTAACACAGC 167
    |||
QY 667 ACTTAATCTTTTACATGCTATGATGCAAAAGCCTTTGTTACATGTTTGTCTGTGAC 726
    |||
DB 166 ACTTAATCTTTTACATGCTATGATGCAAAAGCCTTTGTTACATGTTTGTCTGTGAC 107
    |||
QY 727 CCTTCCCAACAATTTGTTGTGACCTGACACATCCCTCTTGAAGAAACACCAAG 786
    |||
DB 106 CCTTCCCAACAATTTGTTGTGACCTGACACATCCCTCTTGAAGAAACACCAAG 47
    |||
QY 787 ATGATCAGTAATTAAGGGAATCTAGA 815
    |||
DB 46 ATGATCAGTAATTAAGGGAATCTAGA 18
    |||

```

RESULT 3
 BM981862/c 707 bp mRNA linear EST 21-FEB-2003
 LOCUS
 DEFINITION
 UI-CF-EN1-ady-g-12-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
 BM981862
 ACCESSION
 UI-CF-EN1-ady-g-12-0-UI.3, mRNA sequence.
 VERSION
 BM981862.1 GI:19604781
 KEYWORDS
 EST.
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 1 (bases 1 to 707)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL
 Genome Res. 6 (9), 791-806 (1996)
 MEDLINE
 97044477
 PUBMED
 8889548

COMMENT

Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul-mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com) or from Open Biosystems
 (www.openbiosystems.com).
 The following repetitive elements were found in this cDNA
 sequence: 17-707, >L1R5#LTR/Retroviral (matched complement)
 Seq primer: M13 FORWARD
 POLYA=yes.

FEATURES

Location/Qualifiers
 1..707
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-CF-EN1-ady-g-12-0-UI"
 /issue_type="Primary Lung Cystic Fibrosis Epithelial
 Cells"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_1b="UI-CF-EN1"
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site 1: Ecor I; Site 2: Not I;
 UI-CF-EN1 is a normalized cDNA library containing the
 following tissue(s): Primary Lung Cystic Fibrosis
 Epithelial Cells. The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was ligated to an Ecor I adaptor, digested with Not
 I, and cloned directionally into pT73-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dr)18 tail. The
 sequence tag for this library is CTGCTCAGGT.
 TAG-TISSUE=Human Lung Epithelial Cell Lines untreated LPS
 6hr to LPS 24h
 TAG_LIB=UI-CF-EN1
 TAG_SEQ=CTGCTCAGGT"

ORIGIN

Query Match 74.1%; Score 651.2; DB 12; Length 707;
 Best Local Similarity 97.3%; Pred. No. 8.2e-188;
 Matches 673; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

```

QY 151 GAAACATGTGCTGTGCTCACTCAGGTTAAATGATTAAGGCGGTGACAGATGCTT 210
    |||
DB 707 GAAACATGTGCTGTGCTCACTCAGGTTAAATGATTAAGGCGGTGACAGATGCTT 649
    |||
QY 211 GTTAAACAGATGCTTGAAGGACGATCTCTTAAAGTCAATCACTCCCTAATCTCA 270
    |||
DB 648 GTTAAACAGATGCTTGAAGGACGATCTCTTAAAGTCAATCACTCCCTAATCTCA 589
    |||
QY 271 AGTACCAGGACACAAAACCTGCGGAAGGCGGACGAGCCTCTGCTAAGAAAGCCAGG 330
    |||
DB 588 AGTACCAGGACACAAAACCTGCGGAAGGCGGACGAGCCTCTGCTAAGAAAGCCAGG 529
    |||
QY 331 TATTGTCCAAAGTTTCTCCCATGTGATAGCTGAATATATGCTGCGGAGGAAAG 390
    |||
DB 528 TATTGTCCAAAGTTTCTCCCATGTGATAGCTGAATATATGCTGCGGAGGAAAG 469
    |||
QY 391 ACTGACGCTCCCGACCGGACACCCGTAAGGGCTGTGCTGAGAGATTAGTAAA 450
    |||
DB 468 ACTGACGCTCCCGACCGGACACCCGTAAGGGCTGTGCTGAGAGATTAGTAAA 409
    |||

```



```

Db      510 GTAAGCATGAGATGTTATGTATGATATCTAAAGTACAGACTTAATCCTTAC 569
Qy      681 ATTGTCATATGATGAAAGACCTTTGTTTCACATGTTTGTCTGTCACCTCTCCCAACAT 740
Db      570 CTGTCTATATGTGAAAGACCTTTGTTTCACATGTTTGTCTGTCACCTCTCCCAACAT 629
Qy      741 TGTCTTGACACCCGACATCCCTCTTTCGAGAAACACCCACAGATGATCAATAATA 800
Db      630 TGTCTTGACACCCGACATCCCTCTTTCGAGAAACACCCACAGATGATCAATAATA 689
Qy      801 CTAGGAGACTCAGAGCTGGCGGATCTTCATATGCTGAACGCTGGTTCCCGGCTCC 860
Db      690 CTAGGAGACTCAGAGCTGGCGGATCTTCATATGCTGAACGCTGGTTCCCGGCTCC 749
Qy      861 CTTTCTTCTTCTTCTCTATA 879
Db      750 CTTTATTTCTTCTCTATA 768

```

RESULT 8
 B0216036/c 818 bp mRNA linear EST 02-MAY-2002

LOCUS AGENCOURT 7553694 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6046113
 DEFINITION 5', mRNA sequence.

ACCESSION B0216036
 VERSION B0216036.1 GI:20397447

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 818)
 NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: gsa@nci.nih.gov

TISSUE Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LMNL at:

http://image.llnl.gov

Plate: LHM13291 Row: g Column: 10

High quality sequence stop: 577.

Location/Qualifiers

1..818

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6046113"

/tissue_type="embryonal carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_92"

/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NCI;
 Site 2: Sall; Cloned unidirectionally; Oligo-dT primed.
 Average insert size 2.5 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

ORIGIN

Query Match 70.8%; Score 622.4; DB 13; Length 818;

Best Local Similarity 97.5%; Pred. No. 5.8e-179;

Matches 632; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

69 CCAATTGATATGATGAAAGAAATCTTCTGCTTGAATTTCTGTAATCTATGACCT 128

648 CCAATTGATGATGAAAGAAATCTTCTGCTTGAATTTCTGTAATCTATGACCT 589

129 TACCCCAACCCCGTGTCTCTGAAACATGTGCTGCTCAGGGTTAAATGATTA 188

|||||

|||||

Db 588 TACCCCAACCCCGTGTCTCTGAAACATGTGCTCTGCACTGAGATTAAATGATTA 529

Qy 189 AGGGCGGTGAGATGATGCTTTGTTAAACAGATGCTTGAAGGAGAGATGCTTAAAG 248

Db 528 AGGGCGGTGAGATGATGCTTTGTTAAACAGATGCTTGAAGGAGAGATGCTTAAAG 469

Qy 249 TCATACCACTCCCTTAATCTCAATACCCAGGAGACAAAATCGCGAAGGCGCAGG 308

Db 468 TCATACCACTCCCTTAATCTCAATACCCAGGAGACAAAATCGCGAAGGCGCAGG 409

Qy 309 ACTCTGCTTGAAGGAGAGATGCTTGAAGGAGATGCTTGAAGGAGATGCTTGAAG 368

Db 408 ACTCTGCTTGAAGGAGAGATGCTTGAAGGAGATGCTTGAAGGAGATGCTTGAAG 349

Qy 369 ATGGCTCTGAGGAGAGAGATGCTTGAAGGAGATGCTTGAAGGAGATGCTTGAAG 428

Db 348 ATGGCTCTGAGGAGAGAGATGCTTGAAGGAGATGCTTGAAGGAGATGCTTGAAG 289

Qy 429 GTGCTGAGAGAGATGATTAAGAGAGAGATGCTTGAAGGAGATGCTTGAAGGAG 488

Db 288 GTGCTGAGAGAGATGATTAAGAGAGAGATGCTTGAAGGAGATGCTTGAAGGAG 229

Qy 489 GCATCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 548

Db 228 GCATCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 169

Qy 549 GCTCCATCTACTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 608

Db 168 GCTCCATCTACTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 109

Qy 609 CAATACGCTTCTGAG 668

Db 108 CAATACGCTTCTGAG 49

Qy 669 TTAATCCTTACTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 716

Db 48 TTAATCCTTACTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1

RESULT 9

BUI98923 962 bp mRNA linear EST 05-SEP-2002

LOCUS DCEBUE02 DCB Homo sapiens cDNA, mRNA sequence.

DEFINITION BUI98923

ACCESSION BUI98923.1 GI:22717327

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 962)

AUTHORS Xu, X., Gu, J., Liu, F., Qu, J., Zhao, M., Li, Y., Huang, Q., Zhou, J.,
 Song, H., Gu, Y., Yang, Y., Gao, G., Xiao, H., Li, N., Qian, B., Gao, X.,
 Cheng, Z., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M.,
 Lu, G., Cheng, Z., and Han, Z.

TISSUE Homo sapiens cDNA DCB clones

CDNA Library Preparation: (2000)

CDNA Library Arrayed by: Zeguang Han

DNA Sequencing by: Chinese National Human Genome Center at Shanghai

Clone distribution: 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex. 45)

Fax: 86-21-50801922

Email: hanzg@hgc.sh.cn.

Location/Qualifiers

1..962

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/cell_type="dendritic cells"

/dev_stage="mature"

/lab_host="BM25.8"

/clone_lib="DCB"

FEATURES

source

RESULT 11	CD244241/c	991 bp	mRNA	linear	EST 22-MAY-2003
LOCUS	CD244241				
DEFINITION	AGENCOURT_14097645 NIH_MGC_180 Homo sapiens cDNA clone				
ACCESSION	CD244241				
VERSION	CD244241.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	NIH-MGC http://mhc.nci.nih.gov/ .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: csapb@remail.nih.gov Tissue Procurement: Dr. Michael Brownstein cDNA Library Preparation: Invitrogen Corp cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov Plate: NDAM443 row: h column: 24 High quality sequence stop: 728.				
FEATURES	Location/Qualifiers				
Source	1..891 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:30378143" /lab_host="DH10B-Ton A (T1 and T5 phage resistances)" /clone_1lb="NIH_MGC_180" /notes="Organ: Testis; Vector: pCMV-SportE.1; Site.1: NotI site.2: EcoRV (destroyed); Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.68 kb. Library was constructed by (Invitrogen). Note: this is a NIH_MGC Library."				
ORIGIN					
Query Match	69.5%; Score 610.6; DB 14; Length 891;				
Best Local Similarity	93.8%; Pred. No.2.5e-175;				
Matches	682; Conservative 0; Mismatches 34; Indels 11; Gaps 4				
1	1 TGTGGGGAAGCAAGAGAGATGAGATTGTTACTGTGTCTGTGAGAAAGAGTAGACAT	60			
Db	717 TGTGGGGAAGCAAGAGAGATGAGATTGTTACTGTGTCTGTGAGAAAGAGTAGACAT	658			
61	AGGAGACTCCATTTTGTATATGACATCAAGAAAATCTTCTGCTGAGATTCTGTAATC	120			
Db	657 AGGAGACTCCATTTTGTATATGACATCAAGAAAATCTTCTGCTGAGATTCTGTAATC	598			
121	TATGACCTTACCCCAACCCCGTCTCTGTGAACAATGCTGTGTCACTCAGGGTTAA	180			
Db	597 TATGACCTTACCCCAACCCCGTCTCTGTGAACA--GTGCTGTGTCACTCAGGGTTAA	540			
181	ATGATTTAAGGGCGGTGCAGATGTGCTTTGTTAAACAGATGCTTGAAGGACGATGCTC	240			
Db	539 ATGATTTAAGGGCGGTGCAGATGTGCTTTGTTAAACAGATGCTTGAAGGACGATGCTC	480			
241	CTTAAGATCAACACACCTCCCTAATCTCAATACCCAGGACACAAAACATGCGGAAAG	300			
Db	479 GTTAAGATCAACACCAATCCCTAATCTCAATACCCAGGACACAAAACATGCGGAAAG	420			
301	CCGACGGGACCTCTGCTAGAAAGCAGGATTTGTCCACCTTTCTCCCATGTGATAG	360			
Db	419 CCTCAGGACCTCTGCTAGAAAGCAGGATTTGTCCACCAATTTCTCCTCATGGGATAG	360			
361	CCTGAATATGTGGCTCTGTGGGAGGAAACCTGACGCTCCCAAGCCGACACCCGTA	420			

Db	359	TCGTGAATATGCGCTCTGGGAATGAAAGACCTGACCGTCTCCCAACCGAAGCCACTA	300
Qy	421	AAGGCTGTGCTGAGAGGATTAAGTAAAGAGAGAAATGCCTC-TTGCACTTGAGAC	479
Db	299	AAGGCTGTGCTGAGAGGATTAAGTAAAGAGAGAAATGCCTCTTTTGACGTTGAGAC	240
Qy	480	AAGAGAAAGCAATCTGTCTCTCTGCTCTCCCTGGGCAATGAAATGTCGGTATAAACC	539
Db	239	AAGAGAAATGATCTGTCTCTCTGCTC-GTGCTGCGCAATGAAATGTCGGTATAAACC	181
Qy	540	CGATTGTATGCTCCATCTACTAGATAGAGGAAAAACCGCTTAGGGCTGAGTGGAGCC	599
Db	180	CGATTGTATGCTCCATCTACTAGATAGAGGAAAAACCGCTTAGGGCTGAGAGTGAACA	121
Qy	600	TGCGGGAGCAATCTGCTTTGTAAGCACTGAGATGTTATGTGTATGCATATCTAAA	659
Db	120	T-----GCAATCTCTTTGTAAACATGAGATGTTATGTGTATGATATCTAAA	68
Qy	660	GCAAGACACTTAACCTTTACATGTCTGTATGAGCAAGAACCTTTGTATCATGTTGTC	719
Db	67	GCAAGACACTTAATCTTTACCTTTGATGTATGATGCAAAAGACCTTTGTATCATGTTGTC	8
Qy	720	TGCTGAC	726
Db	7	TGCTGAC	1

RESULT 12

CB850840/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

670 bp

mRNA

Linear

EST 22-APR-2003

UI-CF-EN1-ada-a-13-0-UI.s1

UI-CF-EN1 Homo sapiens cDNA clone

UI-CF-EN1-ada-a-13-0-UI.3', mRNA sequence.

CB850840

CB850840.1

GI:3045609

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 670)

Bonardo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene discovery

Genome Res. 6 (9) , 791-806 (1996)

97044477

8889548

Contact: McCray, PB

McCray Lab

University of Iowa

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

Email: paul-mccray@iowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).

The following repetitive elements were found in this cDNA sequence: 18-670, >UTR5#LTR/Retroviral (matched complement)

Seq primer: M13 FORWARD

POLYA=Yes.

Location/Qualifiers

1..670

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-CF-EN1-ada-a-13-0-UI"

/tissue_type="Primary Lung Cyclic Fibrosis Epithelial Cells"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EN1"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-EN1 is a normalized cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT.
TAG_SEQ=None found"

ORIGIN

Query Match 68.1%; Score 598.8; DB 14; Length 670;
Best Local Similarity 95.7%; Pred. No. 9e-172;
Matches 626; Conservative 0; Mismatches 27; Indels 1; Gaps 1;

QY 226 GAAGGAGAGATGCTCTTAAGAGTATACCATCTCCCTAATCTCAAGTACCCAGGACAC 285
DB 670 GAAGGAGAGATGCTCTTAAGAGTATACCATCTCCCTAATCTCAAGTACCCAGGACAC 611
QY 286 AAAAAGCGGAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 345
DB 610 AAAAAGCGGAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 551
QY 346 CTCCTCCATGTATAGCTCTGAATATAGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 405
DB 550 CTCCTCCATGTATAGCTCTGAATATAGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 491
QY 406 AGCCCGACACCCGTAAGAGGTCTGTGCTGAGAGGAGATTAGTAAAGAGAGAGATGCTT 465
DB 490 TGCCCGACACCCGTAAGAGGTCTGTGCTGAGAGGAGATTAGTAAAGAGAGAGATGCTT 431
QY 466 CTGCGAGTTAGACAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 525
DB 430 CTGCGAGTTAGACAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 371
QY 526 CTGCGATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 585
DB 370 CTGCGATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 312
QY 586 CTGAGAGTGGAGCCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 645
DB 311 CTGAGAGTGGAGCAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 252
QY 646 ATGATATCTAAAGACAGACACTTAATCTTATCATTTGATGATGAGCAAGACCTTGG 705
DB 251 ATGATATCTAAAGACAGACACTTAATCTTATCATTTGATGATGAGCAAGACCTTGG 192
QY 706 TTGCATGTTTGTCTGTGAGCCCTCTCCCAATTTCTTGTGAGACCTTGACATCCCTC 765
DB 191 TTGCATGTTTGTGTGTGAGCCCTCTCCCAATTTCTTGTGAGACCTTGACATCCCTC 132
QY 766 CTCTTGAAGAAACCAACAGATGATCAATTAATCTAAGGAACTCAAGAGCTGGCGGG 825
DB 131 CTCTTGAAGAAACCAACAGATGATCAATTAATCTAAGGAACTCAAGAGCTGGCGGG 72
QY 826 ATCTCTCATATGTCGAGACGCTGCTCCCGGTCCTCTTCTTCTCTCTA 879
DB 71 ATCTCTCATATGTCGAGACGCTGCTCCCGGTCCTCTTATTTCTTCTCTTA 18

RESULT 13
B0006641/c 670 bp mRNA linear EST 26-MAR-2002
LOCUS B0006641
DEFINITION UI-H-E11-ayz-b-08-0-UI.s1 NCI_CGAP_E11 Homo sapiens cDNA clone
IMAGE:5845543 3', mRNA sequence.

ACCESSION B0006641 GI:19731541
VERSION B0006641.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 670)
AUTHORS Mammali, Euteria; Primates; Catarrhini; Homnidae; Homo.
TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>
The following repetitive elements were found in this cDNA
sequence: 18-670 >LTR5#LTR/Retroviral (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES

Location/Qualifiers
source 1..670
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5845543"
/issue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="NCI CGAP E11"
/clone_lib="DH10B (Life Technologies)"
/note="Organ: Left Pajvis; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP_E11 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AACTTGCAC.
TAG_TISSUE=chondrosarcoma
TAG_LIB=UI-H-E11
TAG_SEQ=AACTTGCAC"

ORIGIN

Query Match 68.0%; Score 597.6; DB 12; Length 670;
Best Local Similarity 94.6%; Pred. No. 2.1e-171;
Matches 618; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 227 AAGGAGAGATGCTCTTAAGAGTATACCATCTCCCTAATCTCAAGTACCCAGGAGAC 286
DB 670 AAGGAGAGATGCTCTTAAGAGTATACCATCTCCCTAATCTCAAGTACCCAGGAGAC 611
QY 287 AAAAAGCGGAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 346
DB 610 AAAAAGCGGAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 551
QY 347 TCCCATGTGATAGCTGAATATAGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 406
DB 550 TCCCATGTGATAGCTGAATATAGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 491
QY 407 GCCCGACACCCGTAAGAGGTCTGTGCTGAGAGGAGATTAGTAAAGAGGAGGAGGAGGAG 466
DB 490 GCCCGACACCCGTAAGAGGTCTGTGCTGAGAGGAGATTAGTAAAGAGGAGGAGGAGGAG 431

Oy	467	TTGACGTTGAGCAAGAGAAAGGACATCTGTCTCCCGACCTGTCCCTCGGCGCAATGGAATGTC	526
Db	430	TTGACGTTGAGCAAGAGAAAGGACATCTGTCTCTCCGCCCTCTCGGCGCAATGGAATGTC	371
Oy	527	TCGGTATATAAACCCCGATTGTATGCTCCATCTACTGAGATAGGAAAAAACCGCCTTAGAGGC	586
Db	370	TCGGTATATAAACCCCGATTGTATGCTCCATCTACTGAGATAGGAAGAAAAACGCCCTTAGAGGC	311
Oy	587	TGAGAGTGGGACCTTCGCGGCGACCAATACTGCTTTTGTAAAGCACTGAGATGTTTATGTGTGA	646
Db	310	TGAGAGTGGGACATTCGACAGCGACCAATACTGCTTTGTAAAGCACTGAGATGTTTATGTGTGA	251
Oy	647	TGCATATCTAAAAAGCAGACACTTATCCTTTACATATGCTATGATGCAAAAGACCTTTGT	706
Db	250	TGCATATCTAAAAAGCAGACACTTATGTTCTTTACCTTGTCTATGATGCAAAAGACCTTTGT	191
Oy	707	TCACATGTTTTGTCTGTGCTGACCCCTCTCCCCACAATTGCTTGTGACCCCTGACACATCCCCC	766
Db	190	TCACCTGTTTTGTCTGTGACCCCTCTCCCCACATATTGCTTGTGACCATGACACATCCCCC	131
Oy	767	TCTTTCGAGAAACACCCACAGATGATCAAGTAATACTTAAAGGAACTTCAGAGGCTGTGGCGGGA	826
Db	130	TCTTCGAGAAACACCCACGAAATGATCAATTAATACTTAAAGGAACTTCAGAGAGGGCGCGGA	71
Oy	827	TCCTTCATATGCTGAAGCGTGGTTCGCCCGGAGTCCCTTTCTTTCTCTATAT	879
Db	70	TCCTTCATATGCTGAAGCGTGGTTCGCCGAGTCCCTTATTTCTTTCTCTATAT	18

RESULT 14					
BM668693/c	BM668693	613 bp	mRNA	linear	EST 27-FEB-2002
LOCUS	UI-E-CR1-afm-b-10-0-UI .82	UI-E-CR1	Homo sapiens	cDNA clone	
DEFINITION	UI-E-CR1-afm-b-10-0-UI 3', mRNA sequence.				

ACCESSION	BM668693
VERSION	BM668693.1
KEYWORDS	EST.
SOURCE	Homo sapiens (human)

ORGANISM	REFERENCE
<i>Homo sapiens</i>	1
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi	
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
1 (bases 1 to 633)	

AUTHORS Bonaldo, M. F., Lennon, G. and Soares, M. B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE	97044477
PUBMED	8889548
COMMENT	Contact: Soares, MB Coordinated Laboratory for Computational Genomics

University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: Bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregory Hageman
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 18-633, >LTR#LNR/Retroviral (matched complement)
Seq primer: M13 Forward
SOLVA=Yes.

FEATURES	Location/Qualifiers
Source	1. .633

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon: 9606"
/clone="UI-E-CK1-afm-b-10-0-01"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"

```

```

/abc host="VDH10B (Life Technologies) (T1 phage resistant)"
/cle lib="UI-E-CX1"
./note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-CX1 is a normalized cDNA library containing the
following tissue(s): Retina Foveal and Macular. The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GTCC. This library was created for the program,
Gene Discovery in the Visual System, supported by National
Eye Institute (NEI)."
TAG_TISSUE=Foveal and Macular Retina
TAG_LIB=UI-E-CX1
TAG_SEQ=GTCC"

```

Query Match	67.9%;	Score 597;	DB 12;	Length 633;
Best Local Similarity	97.6%;	Pred. No. 3.1e-171;		
Matches 603;	Conservative	0;	Mismatches 15;	Indels 0;
			Gaps	0

ORIGIN

QY	262	CTATCTCAAGTACCCAGGAGACACAAAATCTCGGAAAGGCGCAGAGGACCTCTGCCTAAG	322
Db	633	CTATCTCAAGTACNAGNACACAAAATCTCGGAAAGGCGCAGAGGACCTCTGCCTAAG	573
QY	322	AAAGCCAGGTATTTGTCCAAAGTTTCTTCCCATGTGATAGCTGAATATAGCCCTTGTGGG	381
Db	573	AAAGCCAGGTATTTGTCCAAAGTTTCTTCCCATGTGATAGCTGAATATAGCCCTTGTGGG	514
QY	382	AAGGAAAGACCTGACCCGTCCCCAGCCCGAACCCCTTAAAGGCTCTTGTGAGAGGA	444
Db	513	AAGGAAAGACCTGACCCGTCCCCAGCCCGAACCCCTTAAAGGCTCTTGTGAGAGGA	455
QY	442	TTAGTAAAGAGAGAAATGCCCTCTTGAGTTGAGACAGAGAAAGGACATCTGTCTCT	500
Db	453	TTAGTAAAGAGAGAAATGCCCTCTTGAGTTGAGACAGAGAAAGGACATCTGTCTCT	399
QY	502	GCCGTGCCCTGGGCAATGGAAATGTCTCGGTATPAAAACCGATTGTATGCTCATCTA	561
Db	393	GCCGTGCCCTGGGCAATGGAAATGTCTCGGTATPAAAACCGATTGTATGCTCATCTA	334
QY	562	AGATAGGAAAAACCCCTTTAGGGCTGAGAGGTGGGACCTGCGGGAGACAAATACGCTTTG	622
Db	333	AGATAGGAAAAACCCCTTTAGGGCTGAGAGGTGGGACCTGCGGGAGACAAATACGCTTTG	273
QY	622	TAAAGCATGAGATGTTTATGTGTATGATATCTPAAAAGACAGACACTTAATCCCTTACA	683
Db	273	TAAAGCATGAGATGTTTATGTGTATGATATCTPAAAAGACAGACACTTAATCCCTTACA	214
QY	682	TTGTCTATGATGCAAGACCTTTGTTTCACATGTTTGTCTGCTGACCTCTCCCACAAT	741
Db	213	TTGTCTATGATGCAAGACCTTTGTTTCACATGTTTGTCTGCTGACCTCTCCCACAAT	154
QY	742	GTCCTTGACCCCTGACACATCCCCCTCTTGTGAGAAACACCCACAGATATCATGTAATAC	801
Db	153	GTCCTTGACCCCTGACACATCCCCCTCTTGTGAGAAACACCCACAGATATCATGTAATAC	94
QY	802	TAAAGGAACCTCAGAGCTGGCGGGATCCCTCATATGCTGAAGCGCTGTTCCCGGGTCCC	861
Db	93	TAAAGGAACCTCAGAGCTGGCGGGATCCCTCATATGCTGAAGCGCTGTTCCCGGGTCCC	34
QY	862	CTTCTTCTTCTCTCTATA	879
Db	33	CTTATTTCTTCTCTATA	16
RESULT	15		

CA449778/c 678 bp mRNA linear EST 08-NOV-2002
LOCUS
DEFINITION
UI-H-E11-azd-1-19-0-UI.s1 NCI CGAP E11 Homo sapiens cDNA clone
CA449778
UI-H-E11-azd-1-19-0-UI 3', mRNA sequence.
ACCESSION
CA449778
VERSION
CA449778.1 GI:24814198
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 678)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov
COMMENT
Tissue Procurement: Dr. Jose Mercuende
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soaresuiowa.edu
The following repetitive elements were found in this cDNA
sequence: 16-678, >L1R5#LTR/Retroviral (matched complement)
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES

Source

Location/Qualifiers
1..678
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-E11-azd-1-19-0-UI"
/tissue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_id="NCI CGAP E11"
/note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)
with a modified polylinker; Site 1: Sclor I; Site 2: Not I;
NCI CGAP E11 is a normalized cDNA library containing the
following tissue(s): Chondrosarcoma. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
ACACTGTCAC.
TAG_TISSUE=chondrosarcoma
TAG_L1R=UI-H-E11
TAG_SEQ=ACACTGTCAC"

ORIGIN

Query Match 67.5%; Score 593.4; DB 14; Length 678;
Best Local Similarity 95.1%; Pred. No. 4,1e-170;
Matches 634; Conservative 0; Mismatches 31; Indels 2; Gaps 2;
QY 213 TAAACAGATGCTTGAAGCAGATGCTCTTAAGATCATCAGCATCCCTAATCTCAAG 272
DB 678 TAAACAGATGCTTGAAGCAGCAGCTCTTAAGATCATCAGCATCCCTAATCTCAAG 619
QY 273 TACCCAGGACACAAATCTGCCGAGAGCCGACAGACCTTGTCTAGGAAAGCAGGTA 332
DB 618 TACCCAGGACACGAAATCTGCCGAGAGCCGACAGGCTCTTGTCTAGGAAAGCAGGTA 559
QY 333 TTGTCAAGCTTTCTCCCATGTGATAGCCTGAATATGAGCTGTGAGAAAGGAAAGAC 392
DB 558 TTGTCAAGCTTTCTCCCATGTGATAGCTGAATATGAGCTGTGAGAAAGGAAAGAC 499

QY 393 CTGACCGTCCCCAGAGCCGACACCCGTTAAAGGCTCTGTGTGAGAGATTAGTAAAGA 452
DB 498 CTGACCGTCCCCAGAGCCGACACCCGTTAAAGGCTCTGTGTGAGAGATTAGTAAAGA 439
QY 453 GGAAGAAATGCTCTTGTGACAGTGTGAGACAAAGAGAAAGGACATCTGTCTCTGCTGCTG 512
DB 438 GGAAGAAATGCTCTTGTGACAGTGTGAGACAAAGAGAAAGGACATCTGTCTCTGCTGCTG 379
QY 513 GGCAATGGAATGCTCTGATTAATAAACCCGATTGTATGCTCATCTAAGATAGGAAA 572
DB 378 GGCAATGGAATGCTCTGATTAATAAACCCGATTGTATGCTCATCTAAGATAGGAAA 319
QY 573 AACCCGCTTAGGCTGAGAGTGGAGCTGGGGCAGCAATCTGTTTGAAGCACTGA 632
DB 318 AACCCGCTTAGGCTGAGAGTGGAGCTGGGGCAGCAATCTGTTTGAAGCACTGA 259
QY 633 GATGTTATGTATGATATCTAATAAGCAGACACTTAATCTTATGATGATATGAT 692
DB 258 GATGTTATGTATGATATCTAATAAGCAGACACTTAATCTTATGATGATATGAT 200
QY 693 GCAAGACCTTTGTGACATGTTTGTCTGCTGACCTCTCCGACAAATGTTGTGACC 752
DB 199 GCAAGACCTTTGTGACATGTTTGTCTGCTGACCTCTCCGACAAATGTTGTGACC 140
QY 753 CTGACATCTCCCTCTTGTGAGAAACACCCAGATGATAGTAAATACCTAAGGAACTC 812
DB 139 GTGACATCTCCCTCTTGTGAGAAACACCCAGATGATAGTAAATACCTAAGGAACTC 80
QY 813 AGAGGCTGGCGGATCTCCATATGCTGAACGCTGTTCCCGGGTCCCTCTTCTT 872
DB 79 AGAGGCTGGCGGATCTCCATATGCTGAACGCTGTTCCCGGGTCCCTCTTCTT 21
QY 873 CTCTATA 879
DB 20 CTATATA 14

Search completed: February 28, 2004, 01:07:13
Job time: 3579.96 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2004, 22:08:49 ; Search time 625.802 Seconds

(without alignments)
10389.001 Million cell updates/sec

Title: US-10-016-604-155

Sequence: 1 gagatagggaacacgcct.....atgcaagaccttcttcac 150

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapept 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	150	100.0	150	6	AX594077 Sequence
2	150	100.0	319	6	AX593925 Sequence
3	150	100.0	408	6	AX593926 Sequence
4	150	100.0	879	6	AX593927 Sequence
5	150	100.0	968	6	AX594072 Sequence
6	150	100.0	1505	14	HERVCONF
7	150	100.0	3366	14	HERVCONF
8	150	100.0	8300	14	HERVCONF
9	150	100.0	8708	14	HERVCONF
10	150	100.0	9463	14	HERVCONF
11	150	100.0	9472	14	HERVCONF
12	150	100.0	10029	14	HERVCONF
13	150	100.0	19221	9	AF074086
14	150	100.0	154037	9	AC072054
15	150	100.0	175202	9	AC134684
16	150	100.0	181916	2	AC110927
17	150	100.0	188884	9	AC130367
18	150	100.0	205317	9	AC093856
19	148.4	98.9	521	9	AF276875
20	148.4	98.9	769	9	AF017229
21	148.4	98.9	966	9	AF394944
22	148.4	98.9	1156	9	AK075337
23	148.4	98.9	1877	9	AK098742
24	148.4	98.9	2431	6	BD127902
25	148.4	98.9	2431	9	AK075315
26	148.4	98.9	2517	9	AK090528
27	148.4	98.9	6206	9	AF261945
28	148.4	98.9	8807	14	HEM17833
29	148.4	98.9	9179	9	AF164609
30	148.4	98.9	9181	9	AF164611
31	148.4	98.9	9422	9	AF164615
32	148.4	98.9	9445	9	AF164612
33	148.4	98.9	20540	9	AC010508
34	148.4	98.9	35414	6	BD243619
35	148.4	98.9	35414	9	U73641
36	148.4	98.9	39935	9	AC093231
37	148.4	98.9	41477	9	AC008996
38	148.4	98.9	45092	2	AC130487
39	148.4	98.9	45995	9	AC112702
40	148.4	98.9	59688	2	AC130284
41	148.4	98.9	73221	2	AC022584
42	148.4	98.9	82367	9	AL139421
43	148.4	98.9	91172	9	AC008553
44	148.4	98.9	92171	9	AC004924
45	148.4	98.9	93684	9	AC109513

ALIGNMENTS

RESULT 1	AX594077	AX594077	Sequence 155 from Patent WO0246477.	150 bp	DNA	linear	PAT 13-FEB-2003
LOCUS	AX594077	AX594077	Sequence 155 from Patent WO0246477.	150 bp	DNA	linear	PAT 13-FEB-2003
DEFINITION	AX594077	AX594077	Sequence 155 from Patent WO0246477.	150 bp	DNA	linear	PAT 13-FEB-2003
ACCESSION	AX594077	AX594077	Sequence 155 from Patent WO0246477.	150 bp	DNA	linear	PAT 13-FEB-2003
VERSION	AX594077.1	AX594077.1	GI:28375310	150 bp	DNA	linear	PAT 13-FEB-2003
KEYWORDS	AX594077.1	AX594077.1	GI:28375310	150 bp	DNA	linear	PAT 13-FEB-2003
SOURCE	AX594077.1	AX594077.1	GI:28375310	150 bp	DNA	linear	PAT 13-FEB-2003
ORGANISM	AX594077.1	AX594077.1	GI:28375310	150 bp	DNA	linear	PAT 13-FEB-2003
REFERENCE	AX594077.1	AX594077.1	GI:28375310	150 bp	DNA	linear	PAT 13-FEB-2003
AUTHORS	AX594077.1	AX594077.1	GI:28375310	150 bp	DNA	linear	PAT 13-FEB-2003
TITLE	AX594077.1	AX594077.1	GI:28375310	150 bp	DNA	linear	PAT 13-FEB-2003
JOURNAL	AX594077.1	AX594077.1	GI:28375310	150 bp	DNA	linear	PAT 13-FEB-2003

CHIRON CORPORATION (US)
FEATURES
Location/Qualifiers
source 1..150
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 150; DB 6; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.6e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGATAGGAAAAACCGCCTTAGGGCTGAGGTGGACCTGGCGGACGAATACCTGCTT 60
DB 1 GAGATAGGAAAAACCGCCTTAGGGCTGAGGTGGACCTGGCGGACGAATACCTGCTT 60
QY 61 GTAAGACCTGAGATGTTTATGTATGCAATCTTAAAGACAGCACTTAATCCTTTAC 120
DB 61 GTAAGACCTGAGATGTTTATGTATGCAATCTTAAAGACAGCACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
RESULT 2
AX593925 319 bp DNA linear PAT 13-FEB-2003
LOCUS AX593925
DEFINITION Sequence 3 from Patent WO0246477.
ACCESSION AX593925
VERSION AX593925.1 GI:28375184
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Garcia, P., Hardy, S.F., Williams, L.T. and Escobedo, J.
TITLE Endogenous retroviruses up-regulated in prostate cancer
JOURNAL Patent: WO 0246477-A 3 13-JUN-2002;
CHIRON CORPORATION (US)
FEATURES
Location/Qualifiers
source 1..319
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 150; DB 6; Length 319;
Best Local Similarity 100.0%; Pred. No. 1.6e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGATAGGAAAAACCGCCTTAGGGCTGAGGTGGACCTGGCGGACGAATACCTGCTT 60
DB 1 GAGATAGGAAAAACCGCCTTAGGGCTGAGGTGGACCTGGCGGACGAATACCTGCTT 60
QY 61 GTAAGACCTGAGATGTTTATGTATGCAATCTTAAAGACAGCACTTAATCCTTTAC 120
DB 61 GTAAGACCTGAGATGTTTATGTATGCAATCTTAAAGACAGCACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
RESULT 3
AX593926 408 bp DNA linear PAT 13-FEB-2003
LOCUS AX593926
DEFINITION Sequence 4 from Patent WO0246477.
ACCESSION AX593926
VERSION AX593926.1 GI:28375185
KEYWORDS
SOURCE Homo sapiens (human)

Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Garcia, P., Hardy, S.F., Williams, L.T. and Escobedo, J.
TITLE Endogenous retroviruses up-regulated in prostate cancer
JOURNAL Patent: WO 0246477-A 4 13-JUN-2002;
CHIRON CORPORATION (US)
FEATURES
Location/Qualifiers
source 1..408
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 150; DB 6; Length 408;
Best Local Similarity 100.0%; Pred. No. 1.5e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGATAGGAAAAACCGCCTTAGGGCTGAGGTGGACCTGGCGGACGAATACCTGCTT 60
DB 1 GAGATAGGAAAAACCGCCTTAGGGCTGAGGTGGACCTGGCGGACGAATACCTGCTT 60
QY 61 GTAAGACCTGAGATGTTTATGTATGCAATCTTAAAGACAGCACTTAATCCTTTAC 120
DB 61 GTAAGACCTGAGATGTTTATGTATGCAATCTTAAAGACAGCACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
RESULT 4
AX593927 879 bp DNA linear PAT 13-FEB-2003
LOCUS AX593927
DEFINITION Sequence 5 from Patent WO0246477.
ACCESSION AX593927
VERSION AX593927.1 GI:28375186
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Garcia, P., Hardy, S.F., Williams, L.T. and Escobedo, J.
TITLE Endogenous retroviruses up-regulated in prostate cancer
JOURNAL Patent: WO 0246477-A 5 13-JUN-2002;
CHIRON CORPORATION (US)
FEATURES
Location/Qualifiers
source 1..879
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 150; DB 6; Length 879;
Best Local Similarity 100.0%; Pred. No. 1.5e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGATAGGAAAAACCGCCTTAGGGCTGAGGTGGACCTGGCGGACGAATACCTGCTT 60
DB 561 GAGATAGGAAAAACCGCCTTAGGGCTGAGGTGGACCTGGCGGACGAATACCTGCTT 620
QY 61 GTAAGACCTGAGATGTTTATGTATGCAATCTTAAAGACAGCACTTAATCCTTTAC 120
DB 621 GTAAGACCTGAGATGTTTATGTATGCAATCTTAAAGACAGCACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 681 ATTGCTATGATGCAAGACCTTTGTTTAC 150
RESULT 5

AX594072
 LOCUS AX594072 968 bp DNA linear PAT 13-FEB-2003
 DEFINITION Sequence 150 from Patent WO0246477.
 ACCESSION AX594072
 VERSION AX594072.1 GI:28375305
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 Garcia, P., Hardy, S.F., Williams, L.T. and Escobedo, J.
 Endogenous retroviruses up-regulated in prostate cancer
 Patent: WO 0246477-A 150 13-JUN-2002;
 CHIRON CORPORATION (US)

FEATURES
 source
 1..968
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 100.0%; Score 150; DB 6; Length 968;
 Best Local Similarity 100.0%; Pred. No. 1.5e-35;
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTATGGAAGAAACCCGCTTAGGGCTGAGAGTGGAGCTGGGGGAGCAATACCTGCTT 60
 |||||
 DB 561 GAGTATGGAAGAAACCCGCTTAGGGCTGAGAGTGGAGCTGGGGGAGCAATACCTGCTT 620
 |||||

QY 61 GTAAGACACTGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
 |||||
 DB 621 GTAAGACACTGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 680
 |||||

QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
 |||||
 DB 681 ATTGCTATGATGCAAGACCTTTGTTTAC 710
 |||||

RESULT 6
 HERVCORF 1505 bp mRNA linear VRL 06-APR-1995
 DEFINITION Human endogenous retrovirus mRNA for central open reading frame.
 ACCESSION X82271
 VERSION X82271.1 GI:757869
 KEYWORDS open reading frame.
 SOURCE Human endogenous retrovirus
 ORGANISM Human endogenous retrovirus
 Viruses; Retrovirdae; Retroviridae.

REFERENCE
 1 Lower, R., Tonjes, R.R., Korbacher, C., Kurth, R. and Lower, J.
 Identification of a Rev-related protein by analysis of spliced
 transcripts of the human endogenous retroviruses HTDV/HERV-K
 J. Virol. 69 (1), 141-149 (1995)

JOURNAL
 MEDLINE 95074858
 PUBMED 7983704

REFERENCE
 2 (bases 1 to 1505)
 Toenjes, R.R.
 Direct Submission
 Submitted (20-OCT-1994) R.R. Toenjes, Paul-Ehrlich Inst.,
 Paul-Ehrlich Str. 51-59, P.O.Box 1740, D-63207 Langen, FRG

FEATURES
 source
 1..1505
 /organism="Human endogenous retrovirus"
 /mol_type="mRNA"
 /specific_host="Homo sapiens, patient GH, teratocarcinoma"
 /db_xref="taxon:11827"
 /note="HERV-K"

gene
 1..1505
 /gene="unknown orf"
 255..256
 /gene="unknown orf"
 /note="splice site"

misc_feature
 1..1505
 /note="splice site"

CDS
 273..590
 /gene="unknown orf"
 /codon_start=1
 /protein_id="CAA57722.1"
 /db_xref="GI:757870"
 /db_xref="SPTREMBL:Q69383"
 /translation="MNPSEMQRAKPPRRRRHRNRPPLTHRNKAVTSEQMKLPSTK
 AEPPTWQLKLTQLATKYLENTKVTQTPESMILALMIVSVSAGVNSSEETATIE
 NGP"

misc_feature
 436
 /gene="unknown orf"
 /note="additional nucleotide"

misc_feature
 512
 /gene="unknown orf"
 /note="additional nucleotide"

misc_feature
 533..534
 /gene="unknown orf"
 /note="splice site"

ORIGIN
 Query Match 100.0%; Score 150; DB 14; Length 1505;
 Best Local Similarity 100.0%; Pred. No. 1.5e-35;
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTATGGAAGAAACCCGCTTAGGGCTGAGAGTGGAGCTGGGGGAGCAATACCTGCTT 60
 |||||
 DB 1188 GAGTATGGAAGAAACCCGCTTAGGGCTGAGAGTGGAGCTGGGGGAGCAATACCTGCTT 1247
 |||||

QY 61 GTAAGACACTGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
 |||||
 DB 1248 GTAAGACACTGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 1307
 |||||

QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
 |||||
 DB 1308 ATTGCTATGATGCAAGACCTTTGTTTAC 1337
 |||||

RESULT 7
 HERVENV 3366 bp mRNA linear VRL 30-MAR-1995
 LOCUS
 DEFINITION Human endogenous retrovirus env mRNA.
 ACCESSION X82272
 VERSION X82272.1 GI:757871
 KEYWORDS env gene.
 SOURCE Human endogenous retrovirus
 ORGANISM Human endogenous retrovirus
 Viruses; Retrovirdae; Retroviridae.

REFERENCE
 1 Lower, R., Tonjes, R.R., Korbacher, C., Kurth, R. and Lower, J.
 Identification of a Rev-related protein by analysis of spliced
 transcripts of the human endogenous retroviruses HTDV/HERV-K
 J. Virol. 69 (1), 141-149 (1995)

JOURNAL
 MEDLINE 95074858
 PUBMED 7983704

REFERENCE
 2 (bases 1 to 3366)
 Toenjes, R.R.
 Direct Submission
 Submitted (20-OCT-1994) R.R. Toenjes, Paul-Ehrlich Inst.,
 Paul-Ehrlich Str. 51-59, P.O.Box 1740, D-63207 Langen, FRG

FEATURES
 source
 1..3366
 /organism="Human endogenous retrovirus"
 /mol_type="mRNA"
 /specific_host="Homo sapiens, patient GH, teratocarcinoma"
 /db_xref="taxon:11827"
 /note="HERV-K"

gene
 1..3366
 /gene="env"
 437..2536
 /gene="env"
 /codon_start=1
 /protein_id="CAA57723.1"
 /db_xref="GI:757872"

CDS
 437..2536
 /gene="env"
 /codon_start=1
 /protein_id="CAA57723.1"
 /db_xref="GI:757872"

```

/db_xref="GOA:O69384"
/citation="WNSBEMOKARPRRRRRHRNAPLTHKNKRVTSBEOKLEPSTYKK
AEPPTMAOLKLTQLATKYLETNTVTVTPRESMLLAALTVSMVSLPMPDAAANVTY
YMAVAPFPLIRAVTWMNDPTEVYVNSVWPGPIDRCPAKPEEGEMINI SIGEYH
PPIICGRACGLMPAVQMWLEVPVSPICFPTVHMSGSLRPRNVLODFVSORSL
KFRKGRGPCPEIKPRESKENTVLWEECVANSAVILNBERGTLIDNAPGQPFHNSL
GQTOSCBQAVSPAVDSLTSLDKHKKQLQSYPYMEWGKIGSTREPKTVSVSGPR
EHPMLRLTVASHHIRLWSGQTLERTRKPFVITIDNLSLTVLQCVKPEYMLVVG
NIVLKPDSQITTCENCRLLTCTIDSTFNWQHRILVVRAGEVILVSDRPEASVH
ILTVLKGVLNRSRFLFTLLAVIMGLAVATAVAVGVALHSSQSVNFVMDQKNS
TRLNOSSIDOKLANQINDROTVIMWGDLMSEHRFOLCCMNTSDFCITPOIYN
ESEHHMVMVRHLOGREDNLITLIDSKLEQFEASKAHILNPGTEALAGAGDLANL
NPYVWKTIGSTITINLILIVCLFLCLLVRCCTQQLRRSDRERAMTVAVLSKSK
GQNGKSKRDQIVTVSV"

misc_feature
600
    /gene="env"
    /note="additional nucleotide"
676
    /gene="env"
    /note="additional nucleotide"

ORIGIN
Query Match      100.0%; Score 150; DB 14; Length 3366;
Beet Local Similarity 100.0%; Pred. No. 1.4e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGATAGGGAACAAACCCGCTTAGGAGCTGGAAGTGGAACCTGCGGAGCAATACCTGCTT 60
Db 3051 GAGATAGGGAACAAACCCGCTTAGGAGCTGGAAGTGGAACCTGCGGAGCAATACCTGCTT 3110
Qy 61 GTAAAGCCTGAGATGTTTATGTGTATGCATATCTTAAAGACAGACTTAATCCTTTAC 120
Db 3111 GTAAAGCCTGAGATGTTTATGTGTATGCATATCTTAAAGACAGACTTAATCCTTTAC 3170
Qy 121 ATTGCTATGATGCAAGAACCCTTGTTGTCAC 150
Db 3171 ATTGCTATGATGCAAGAACCCTTGTTGTCAC 3200

RESULT 8
LOCUS HEN17834 8300 bp DNA linear VRL 21-OCT-1999
DEFINITION Human endogenous retrovirus K (HERV-K) elements, clone SD-C7-34LTR.
ACCESSION Y17834
VERSION Y17834.1 GI:4185945
KEYWORDS env gene; env protein; gag gene; gag protein; long terminal repeat;
SOURCE Human endogenous retrovirus K
ORANISM Viruses; Retroid viruses; Retroviridae.
REFERENCE
1 Tonjes,R.R., Czanderne,F. and Kutth,R.
AUTHORS Genome-wide screening, cloning, chromosomal assignment, and
TITLE expression of full-length human endogenous retrovirus type K
JOURNAL J. Virol. 73 (11), 9187-9195 (1999)
MEDLINE 99445825
PUBMED 10516026
2 (bases 1 to 8300)
REFERENCE
1 Toenjes,R.R.
AUTHORS Direct Submision
TITLE Submitted (22-JUL-1998) R.R. Toenjes, Paul-Ehrlich Inst.,
JOURNAL Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-63207 Langen, FRG
COMMENT Related sequence Y10390.
FEATURES
Location/Qualifiers
1..8300
    /organism="Human endogenous retrovirus K"
    /proviral
    /mol_type="genomic DNA"
    /specific_host="human"
    /db_xref="taxon:45617"
    /clone="SD-C7-34LTR"
    /note="assigned to human chromosome 7"

```

gene
CDS
'
gene
CDS
gene
CDS
gene
CDS
LTR
/note="3' LTR"

```
32. . 2032
/gene="gag"
32. . 2032
/gene="gag"
/codon_start=1
/product="gag protein"
/protein_id="CA76884.1"
/db_xref="GI:4185946"
/db_xref="GOA:Q9YNA6"
/cramblation="MGOTGSKIKSVKASVSYFIKILKRGVVKVSTNLIKLFQILEQ
FCMPPEGGTIDLKMKRIGKELKQAKRKNIIPLVTMNMWAIKALPEOFEEEDV
SVNAPSGCIIDCENMTRKKSOKETGEHCEYVAPRMASTONPVNOLOEVYPT
LKLBKSGELVYCNESBESKRGTSIPACQVPTTLPOQYKVENTKOPPEVATKREB
QYRPPEQYQYQGMPPAPQGRAPYQPPPTRRILTPAPSRQSKLHIIIDKSEED
TEAOWPFTLPEMPPEGQAQGEPPTEARVKSFSIKLDMKEGVQYQPNPSYMT
LIDSAJHHRILPYDMEILAKSSLSPSQFLQFTKWIIDGOEVRNRAANPUNIDA
DOLIGIGQWMSITISQALMONBAIEYRAICLAMEKIIDPGSTCFSEFNVRQSEK
YPPDPAIQQDPAQKSIDADERAKIYIELMAENANPCSGAIPKIKVAGSDVSE
YVKADGIGGEMKAMLMQAQITGVVLGGVRTFRKCTNCGQIGHLKNCPLANKON
ITIDATTTGRRPDLPCRCKKGMASQCSKBDKNGQPLSGNEGRQDPAPOQTGA"
PIOPVPGFQGGQPPLSOVFOGISOLPQYNNCPPOAAVQ"
3051. . 5669
/gene="pol"
3051. . 5669
/gene="pol"
/codon_start=1
/product="pol protein"
/protein_id="CA76884.1"
/db_xref="GI:4185947"
/db_xref="GOA:Q9MJR4"
/db_xref="SPRTEMBL:Q9MJR4"
/cramblation="MLTDLRAVNAVIOPMGELDQPLSPMSPIPKDWPLIIDKDFE
TIPLAEOCEKFAFTIPIANNKEBAPTPOKWLPOGMLNSPTICOTFVGHALPVREK
FSDCIHICIDILCAEATKDKLIDCYTPOAEVANAAGLAISDKIQTSPHYLGMQ
IENKRIQKIDKILDKTLINDROKLGIDIMWTRPLTGLPYAMSNLPSLIGDSDL
NSKMLTPREAKELIYBEKIQSQNRIIDPRLAQLILATASPGITIIONTDVE
WSPLSHSVKPTIYLDQIATLIGOTSLRIKLCGNPDKI VAPLVEQVRQAFINS
AMKIGAPVGIIDNHYPKTIPOFLKTLTWIIPKIRREPLEALTVFPGSSNGGA
AYTPKREKVIITPYQSAQRAELVAVITVLDPDPPIIISDAVVOATDEITALIK
YSMDLODNLNLIOOTVRKRNPFYITHI RAHNLGPRLTKANEGADLVSSALIK
OELALHVNAAGLKKEEDVYKQAKDIOCHOCQOYLH.PIQEGANPBGCLPNNALM
QMDTVHSEKRLSYHVTVDYTSHTIWAITQIBESISHYKHLISCFYAMGPBEKTK
ITDNGVYSKAFQKFLSQWKISHTTGISYNSQQAIVERNRTLKTQLVKQKGGDSK
ECTTPOMQNLATLYNLFINYNQUTTSABOHLTGKNSPHEBKLIMNDONKITE
IGKATYWRGACVSPGNOLPWIIPTRHLKFYNEPIRDAKKSTSAETETSSQSYDS
ODEBNGVRRRDEVAIHIOEGRAANLGTTKEADANYSKISHEKGDTPRERYACSIDD
CINGKSYARSSCS"
```


ORIGIN

Query Match 100.0%; Score 150; DB 14; Length 8300;
Best Local Similarity 100.0%; Pred. No. 1.4e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGATAGGAAAACCCGCTTAGGGCTGAGAGTGGAGACCTCGGGCAGCAATACCTGCTT 60
Db 7985 GAGATAGGAAAACCCGCTTAGGGCTGAGAGTGGAGACCTCGGGCAGCAATACCTGCTT 8044

Qy 61 GTAAGACCTGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
Db 8045 GTAAGACCTGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 8104

Qy 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
Db 8105 ATTGCTATGATGCAAGACCTTTGTTTAC 8134

RESULT 9
AF490464
LOCUS AF490464 8708 bp DNA linear PRI 08-APR-2002

DEFINITION Homo sapiens HERV-K' long terminal repeat, complete sequence; and
envelope glycoprotein gene, partial cds.
AF490464
VERSION AF490464.1 GI:20067082

KEYWORDS
SOURCE Homo sapiens (human)

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Reus, K., Stuhr, T., Mayer, J. and Meese, E. U.
TITLE Haplotype analysis of Human Endogenous Retroviruses: A genetic
variant of HERV-K(HML-2.HOM) with an intact YXDD motif of reverse
transcriptase
Unpublished

JOURNAL
REFERENCE 2 (bases 1 to 8708)
AUTHORS Reus, K., Stuhr, T., Mayer, J. and Meese, E. U.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-2002) Humangenetik, Universitaet des Saarlandes,
Kirtbergstrasse, Homburg 66421, Germany

FEATURES
source
1..8708
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/note="from Asian sample
endogenous virus: HERV-K'"
239..8708
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/note="endogenous virus: HERV-K'; from Asian sample"

LTR
misc_feature
misc_feature
misc_feature
CDS
6688..>8708
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AA095525.1"
/db_xref="GI:20067083"
/translation="MNPEMQRKAPRRRRHRNRAPLTHKNKAVTSEBQKLPSTKK
AEPPTMAQLKLTQLATYLTENTVTQTPEBMLAALMIIVMVVSLPMPAGAAANTY
YMAVPPPLIRAVTMDNPEVAVVNSVWPGPIDRCRPAKPEEGMNIISGYHY
PPICLGAPGCLMPAVONMVLVEPTVSPICFTYHMGSLRPRVNYLQDFSYRSL
KFRKGRPCREIKPKSKNTEVLVMEBCVANSVILQNNRFTIIDNAPRQFHNCS

ORIGIN

Query Match 100.0%; Score 150; DB 9; Length 8708;
Best Local Similarity 100.0%; Pred. No. 1.4e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGATAGGAAAACCCGCTTAGGGCTGAGAGTGGAGACCTCGGGCAGCAATACCTGCTT 60
Db 799 GAGATAGGAAAACCCGCTTAGGGCTGAGAGTGGAGACCTCGGGCAGCAATACCTGCTT 858

Qy 61 GTAAGACCTGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
Db 859 GTAAGACCTGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 918

Qy 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
Db 919 ATTGCTATGATGCAAGACCTTTGTTTAC 948

RESULT 10
AY037929
LOCUS AY037929 9463 bp DNA linear VRL 29-OCT-2001

DEFINITION Human endogenous retrovirus K115 complete genome.
AY037929
VERSION AY037929.1 GI:16507983

KEYWORDS
SOURCE Human endogenous retrovirus K115
Human endogenous retrovirus K115
Viruses; Retroid viruses; Retroviridae.

REFERENCE
AUTHORS Turner, G., Barbulessu, M., Su, M., Jensen-Seaman, M. I., Kidd, K. K. and
Lenz, J.
TITLE Insertional polymorphisms of full-length endogenous retroviruses in
humans
JOURNAL Curr. Biol. 11 (19), 1531-1535 (2001)

JOURNAL
MEDLINE 21475517
PUBMED 11591322
REFERENCE 2 (bases 1 to 9463)
AUTHORS Turner, G., Barbulessu, M., Su, M., Jensen-Seaman, M. I., Kidd, K. K. and
Lenz, J.
TITLE Direct Submission
JOURNAL Submitted (04-JUN-2001) Molecular Genetics, Albert Einstein College
of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA

FEATURES
source
1..9463
/organism="Human endogenous retrovirus K115"
/proviral
/mol_type="genomic DNA"
/specific_host="Homo sapiens"
/db_xref="taxon:166123"
1..960
1104..6740
/note="gag-pro-pol; two -1 frameshifts predicted to occur
to produce a fusion protein; the location of frameshifts
has not been determined"

LTR
misc_feature
misc_feature
CDS
6442..8541
/codon_start=1
/product="putative env"
/protein_id="AAL18259.1"
/db_xref="GI:16507984"
/translation="MNPEMQRKAPRRRRHRNRAPLTHKNKAVTSEBQKLPSTKK
AEPPTMAQLKLTQLATYLTENTVTQTPEBMLAALMIIVMVVSLPMPAGAAANTY
YMAVPPPLIRAVTMDNPEVAVVNSVWPGPIDRCRPAKPEEGMNIISGYHY
PPICLGAPGCLMPAVONMVLVEPTVSPICFTYHMGSLRPRVNYLQDFSYRSL
KFRKGRPCREIKPKSKNTEVLVMEBCVANSVILQNNRFTIIDNAPRQFHNCS

GOTOSCPASPAVSDLTRESLDKHKHKLQSPYPMWEGEKRLSTPRKTVSPVSG
EHEBELWRLTVAASHHIRIIRMSGNQTLFETDRKPFYTVLNSLITPLQSCVAPRYMLVVG
NIVIKPDSQITTCENCLRLTICIDSTENWQHRILLVREBVMILPVSDRWESPSVH
ILTEVLKGVNRSKRPIFTLIAYMGLIATVAAGVALHSSVQSVNVNDGQKNS
TRLNWSQSSIDOKLANQINDLRQTVIMWGRILMSLEHRFQLOCDMVNTSDICITPQIYN
TSEHHMDWVRRLHOGREDNLTLDISKLKEQIFASAKHANLIVGTEAIAVAGLANL
NPVTWKTIGSTTILNLIILVCLFCLLLVCRCTQQLRDSHREBRAMMTMAVLSKRX
GNVGVSKRQOITVSV"

ORIGIN LTR 8496. .9463

Query Match 100.0%; Score 150; DB 14; Length 9463;
Best Local Similarity 100.0%; Pred. No. 1.4e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAAAGCCGCTTAGGCTGAGAGTGGGACCTGCGGGCAGCAATACCTGCTT 60
DB 553 GAGATAGGAAAAAAGCCGCTTAGGCTGAGAGTGGGACCTGCGGGCAGCAATACCTGCTT 612

QY 61 GTAAGACACGAGATGTTATGTGTATGATATCTTAAAGACAGCACTTAATCCTTTAC 120
DB 613 GTAAGACACGAGATGTTATGTGTATGATATCTTAAAGACAGCACTTAATCCTTTAC 672

QY 121 ATTGCTATGATGCAAGACCTTTGTTAC 150
DB 673 ATTGCTATGATGCAAGACCTTTGTTAC 702

RESULT 11
AF164614
LOCUS AF164614 9472 bp DNA linear PRI 03-SEP-1999
DEFINITION Homo sapiens endogenous retrovirus HERV-K108, complete sequence.
ACCESSION AF164614
VERSION AF164614.1 GI:5802820
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 9472)
Barbulescu, M., Turner, G., Seaman, M.I., Deinard, A.S., Kidd, K.K. and
Lenz, J.
TITLE Many human endogenous retrovirus K (HERV-K) proviruses are unique
to humans
JOURNAL Curr. Biol. 9 (16), 861-868 (1999)
MEDLINE 99400989
PUBMED 10469592
REFERENCE 2 (bases 1 to 9472)
Barbulescu, M., Turner, G., Seaman, M.I., Deinard, A.S., Kidd, K.K. and
Lenz, J.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-1999) Molecular Genetics, Albert Einstein College
of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA
FEATURES
SOURCE Location/Qualifiers
1. .9472
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1. .9472
/rpt_family="human endogenous retrovirus HERV-K108"
1. .968
1112. .6749
/gene="gag-pro-pol"
join(1112. .3108,3108. .3915,3915. .6749)
/gene="gag-pro-pol"
/note="-1 frameshifts presumed to occur within the codons
for the last amino acids in the gag and pro open reading
frames"
/codon_start=1
/product="Gag-Pro-Pol protein"
/protein_id="AAS1797.1"
/db_xref="GI:5802821"
/translation="MGQTKSKIKSVASYSLFIKILKRGVAVKSTNKLILKPOLIEQ

FCWPPEQGTLDLQKMKRIGKELKQAGKNGINIPLTWMDMALIKALPEPOTEBDSV
SVSDAGSGCITDGNENTRRKKSQKETEGLHCEIYABEVMWMSOTNDYNOLOQVITYPT
LKIEGPELVEPSESKEPRGTPLPAGQVFLYIQPKQVKNENTQPPVAQVPPAE
QVRRPESQVLEPMPGAPQAPVPPVRRPLRNLNAPPSQKLEHIIIDSKRKG
TEWQPPVLEPMPGEGAOEGEPVPEAPKPIPLTWKTEKPVWQVMPPLPKLEL
HLIANLEKRGHIEPSPWNSPVFVIOKSGKMRMLTDLRANAVIQMPLPGPIL
SPAMIKQMDPLIITDKCEFTIPLAEOCEKAFTIPIANNEPATRPFQWKLPOQM
INSPTICQFVBALQPVREKSPDVIITHCIDDIICAEFTKDLIDCYPLOAEVANA
GLAIAADKIQTSPTFFHLMQOIBNRKIKQKIEIRDTLKTLDNFQKLGEDINWIRP
LGIPYAMSNLFEIILRGSDNLNSKRLNTEPATEIKLVEBKIQSAQINRIDLPAQLD
LAFATASPTGLIIONTDVWESFLSHSVKTEFLYLDQATLIGQTRRLIKLQND
PDKIVPLFTEQVRQAFINSKAMKIGLANFVGIIDNHVPTKIFQELKLTWTLPIKT
RREPLNALTVEFDGSSNGKAAVTGPKERYIKTPYQSAQAEVAVITVLDQDPQIN
IISDAVVAQATRDVETALIKYSMDOLNQLFLLQOTYRKRRPFTYTHIRAHNTLP
GPITRANBQDLVLSALIKAGELHLYTHNMAGLKMKPDYTHKQADIVQCTQCY
LHLPTQEAGVNERGLCPNALMOMDVTHVSPGRSLYVHTVDTYSHFTAKTCGTGBST
SHVKKHLSCFAWGVPERIKITDNGPYCSKAKQKLSQKISHTTGIIPNSQGAIV
ERTNRILKTQVKEKGDSEKCTIPOMOLNALLNLINFTLIRNQTTSAEQHLNCK
KNSPHEGKLIWMDKNGKTBWEIGKVTGKGVCSVGENQULPVMIPTRLIKYNEBI
RDAKSKTSAETENOSSTVYDODEONGDVRDREVAIHQEGRAANIGTTMEADVSYK
ISREHKDTPNRYAACSIDDCINGAKSPACSSCS"

gene 6451. .8550
/gene="env"
/codon_start=1
/product="envelope protein"
/protein_id="AAS1798.1"
/db_xref="GI:5802822"

ORIGIN LTR 8504. .9472

Query Match 100.0%; Score 150; DB 9; Length 9472;
Best Local Similarity 100.0%; Pred. No. 1.4e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAAAGCCGCTTAGGCTGAGAGTGGGACCTGCGGGCAGCAATACCTGCTT 60
DB 561 GAGATAGGAAAAAAGCCGCTTAGGCTGAGAGTGGGACCTGCGGGCAGCAATACCTGCTT 620

QY 61 GTAAGACACGAGATGTTATGTGTATGATATCTTAAAGACAGCACTTAATCCTTTAC 120
DB 621 GTAAGACACGAGATGTTATGTGTATGATATCTTAAAGACAGCACTTAATCCTTTAC 680

QY 121 ATTGCTATGATGCAAGACCTTTGTTAC 150
DB 681 ATTGCTATGATGCAAGACCTTTGTTAC 710

RESULT 12
LOCUS HEN17832 10029 bp DNA linear VRL 21-OCT-1999
DEFINITION Human endogenous retrovirus K (HERV-K) elements, clone C7.
ACCESSION Y17832
VERSION Y17832.2 GI:4581240
KEYWORDS env gene; env protein; gag gene; gag protein; long terminal repeat;
pol gene; pol protein.
SOURCE Human endogenous retrovirus K
ORGANISM Human endogenous retrovirus K
VIRUSES; Retroviral viruses; Retroviridae.
REFERENCE 1
AUTHORS Tonjes, R.R., Czauderna, F. and Kurth, R.
TITLE Genome-wide screening, cloning, chromosomal assignment, and
expression of full-length human endogenous retrovirus type K
JOURNAL J. Virol. 73 (11), 9187-9195 (1999)
MEDLINE 99445825
PubMed 10516026
REFERENCE 2
AUTHORS Tonjes, R.R.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-1998) R.R. Tonjes, Paul-Ehrlich Inst.,
Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-63207 Langen, FRG
REMARK 3 (bases 1 to 10029)
REFERENCE 3
AUTHORS Tonjes, R.R.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1999) R.R. Tonjes, Paul-Ehrlich Inst.,
Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-63207 Langen, FRG
COMMENT On Apr 12, 1999 this sequence version replaced gi:4185937.
Related sequence Y10390.
FEATURES
source location/Qualifiers
1..10029
/organism="Human endogenous retrovirus K"
/proviral
/mol_type="genomic DNA"
/specific_host="human"
/db_xref="taxon:45617"
/clone="C7"
/clone_lib="PI (Genome Systems, St. Louis, MO, USA)"
/notes="assigned to human chromosome 7"
1..83
/rpt_family="Alu"
243..426
/rpt_family="Alu"
427..1394
/note="5' LTR"
1397..1414
1538..3538
/gene="gag"
1538..3538
/gene="gag"
/gene="gag"
/codon_start=1
/product="gag protein"
/protein_id="CA476878.1"
/db_xref="GI:4185938"
/db_xref="GOA:O9YNA9"
/db_xref="SPTREMBL:O9YNA9"
/translation="MGQTSKIKSKVSYSLFIKILKRGVKNVSTNLKLPQIIQ
FCWPFQGTLDLDMKRIKELKQAGRKNIIPVTWMAIIKALIEPQEDSEV
SVSDAPSCIIIDNENTRKISQKETEGLKYEYVPAWMASTONVDNOLQEVYPT
LKLEGGKPELVGSESKDRTSLPAGOVPTLOPOKQVKNSTOPVAVOYMPRAE
QYRRPESQYGRGMPAPRABAPYPPPTRRRLNTPRPSQSKHEITDKSRKED
TEAMPVPTLEPMPGEGADGEPPTVEARKSFISKLDMKEGVKQYSPNSYMT
LDSIAHGRLIPYDWEILAKSLSPSQFLQFTKMTIDGOVRRNRANPPNIDA
DOLIGQNMSTISQALQNEALIEQVRAICLRAEMKIDPGSTCPSPFNQSKSEY
YPPFARLOVAVAKSIDAKARVIVELMAYENNPCCOSAIKPLKGVAGSVISE
YVACDGIIGAMKAMLMAGALITGVIVGQVTRGRCVNCGOIGHKKRCPLNKN
ITIOATTTGRRPPLCPCKKKKMAQCSKPKKNOPIPSGNEORQOPAPOOTGAF
PIQFPVQGFQGOOPPLSOVFOGISQLPOTNCPPOAAVQO"
gene
4557..7175
/gene="pol"
CDS
4557..7175

/gene="pol"
/codon_start=1
/product="pol protein"
/protein_id="CA476879.1"
/db_xref="GI:4185939"
/db_xref="GOA:O9YNA9"
/db_xref="SPTREMBL:O9YNA9"
/translation="MLTDLAVANAVIQMPGLPGLSPAMI PKDMPLIITDLKCFE
TIPLAEDCKEPAFTIPAINNKEPATFOKVLIPQGLNSPTICQTFVGAALQVREK
PSCVITHCIDDLICAEFTDKLIDCTPTQABYANAGLIAASKIOTSPRHFGMO
IENRKIPQKIEIRKDTLKTUNDFOKLGDINMIRPLGIPITAMSLFSLRDSPL
NSKMLTPPEATKEIKLVEEKIQAQINRIIDPLKGNPDKIIVPLTEQVROAFINS
WSPKSTVTFTLVDIATILGQTRILIKLCSNDPDIIVPLTEQVROAFINS
AMKIGLANPFIIDNHYPKTKI FOELRLITWILPEKTRREPLEVALTFVDESGNKA
AYGKREKRVITKTPYOSARAEIVAVITVLDPOPINIISDAVAVOATDVERALK
YSMDQNLQNFILQTVRRKNPFYITTHRAITNLGPIITKNEQADLVSSALITK
QELHALTHVNAAGLKNFVDVTKQAKDVOHCTQOVHLPTDAGVNPGLCPNALM
QMDVTHVSFGRSLSYVAVTDYSHFTWATCQGBESTSHVKHLSCFVWGVPEKIK
TDNPGYCSKAFQKFLSQMKISHTTGIPYNSQGAIVERNTRILKTOLVQKEGDEK
ECTTPQOMNLALYTLNFINIYRNQTTSAEHLTGKNSPHEGLIMNDSKXKWE
IGKVIITGRGPAVCSPEGNOLPWIPTRHLEFVNEPIRDAKKSATETSSQSTVS
ODEGNGVRRTEDEVAITQEBRANLTGTTKADAVYSIKSEHKDTPPREYACSLD
CINCKSPYACRSCS"
6892..8976
/gene="env"
6892..8976
/gene="env"
/codon_start=1
/product="env protein"
/protein_id="CA476880.1"
/db_xref="GI:4185940"
/db_xref="GOA:O9YNA9"
/db_xref="SPTREMBL:O9YNA9"
/translation="WQKAPRRRRRRNRAPLTHKNNKNTSEBOMKLBSTKAEPT
WAQKLTQATLAXYLENTKVTQTPESMLALAMIVSNVSLPMPAGAAANYTWAYV
PPEPLIRAVTMDNPTVEYVNDVWVGEPIDIDCPSABEESGMINISYHVPICL
GRAGCLMPAVOMLVEPVPSPICRTYHMGSPAPRVNLYODYSRLKFRK
KDCPKXEPKESKNTVELVMEBCVANAVALIONNEFETIIDMAPRGQFYNCSGQTS
COSAOVSPAYDSLTSLDGCHKKULQSPFPMGEKGISTPPEKYSPIGSPHPEL
WRUTVASHIRINSNOTLETDRKPEPTIDLNSSLVPLQSCVPEPMYLVGNIVIK
PDSQTTTCENCRILTICIDSTFNMQHRLLVAREGWI PVSMRPMWASPSVHLTEV
LKGVNLSKRFIFTLIAVINGLIAVATAVAGVALSSVQSVNFDVMDKNSRLMN
SOSISIDQKLANQINDLRQIVIMMGDRMSLEHFOLOCDNNTSDFCITPOIYSEBH
MDWVRHLQGRNLTLDISKKEQIFEASKAHNLVPGTEALAGVADGLANIPVW
VKTIQSTTINILILVCLFCLLVCRCTQQLRDSDBRERAMTNAVLSKRGANG
KSKRDQIVTVS"
8931..9898
/note="3' LTR"
9899..9903
/rpt_family="Alu"
ORIGIN
Query_Match 100.0%; Score 150; DB 14; Length 10029;
Best Local Similarity 100.0%; Pred. No. 1.4e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGTAGGAAAAACCCCTTAGGCTGAGAGTGGAGCTCGCGGCGAGCAATACCTTT 60
DB 9491 GAGATGAGGAAAAACCCCTTAGGCTGAGAGTGGAGCTCGCGGCGAGCAATACCTTT 9550
QY 61 GTAAGACGAGAGTGTGTTATGATGATCTAATAAGACAGCACTTAATCCTTAC 120
DB 9551 GTAAGACGAGAGTGTGTTATGATGATCTAATAAGACAGCACTTAATCCTTAC 9610
QY 121 ATTGCTATGATGAAGAAGACCTTTGTTTAC 150
DB 9611 ATTGCTATGATGAAGAAGACCTTTGTTTAC 9640
RESULT 13
LOCUS AF074086 19221 bp DNA linear PRI 11-MAR-2002
DEFINITION Homo sapiens tandemly repeated human endogenous retrovirus HERV-K

(HML-2.HOM), complete sequence.
ACCESSION AF074086
VERSION AF074086.2 GI:9558700
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 19221)
Meyer, J., Sauter, M., Racz, A., Scherer, D., Mueller-Lantzsch, N. and Meese, E.
TITLE An almost-intact human endogenous retrovirus K on human chromosome 7
JOURNAL Nat. Genet. 21 (3), 257-258 (1999)
MEDLINE 99178257
PUBMED 10080172
REFERENCE 2 (bases 1 to 19221)
AUTHORS Reus, K., Mayer, J., Sauter, M., Scherer, D., Muller-Lantzsch, N. and Meese, E.
TITLE Genomic organization of the human endogenous retrovirus HENV-K (HML-2.HOM) (ERV-K) on chromosome 7
JOURNAL Genomics 72 (3), 314-320 (2001)
MEDLINE 21295053
PUBMED 11401447
REFERENCE 3 (bases 1 to 19221)
AUTHORS Mayer, J., Mueller-Lantzsch, N. and Meese, E. U.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-1998) Institut fuer Humangenetik, Universitaet des Saarlandes, Homburg 66421, Germany
REFERENCE 4 (bases 1 to 19221)
AUTHORS Reus, K., Mayer, J. and Meese, E. U.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-2000) Institut fuer Humangenetik, Universitaet des Saarlandes, Homburg 66421, Germany
REMARK Sequence update by submitter
COMMENT On Jul 26, 2000 this sequence version replaced gi:4456987.
FEATURES
source
1. 19221
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/clone="Y20-B19"
/clone_id="Lawrence Livermore National Laboratory Single Chromosome Library: L07NCO1"
1. 17976
/note="HML-2.HOM; tandemly repeated"
/rpt_family="Human endogenous retrovirus HENV-K"
/rpt_type="dispersed"
1. 968
971. 988
/note="lysine primer binding site"
1112. 3112
/gene="gag"
1112. 3112
/gene="gag"
/codon_start=1
/product="gag protein"
/protein_id="AAF88165.1"
/db_xref="GI:9558702"
/translation="MGQTSKISKVASYLUSFIKILKRGVAVSTKNLIKLFQIIEQ
FCMPPEQGLTDLKDKRIKELKQAGKRNIIPLTYNMWAIKALFEPQTEBSV
SVSDPSCIIIDNENTRKKSQETEGHCEYVAPMAOSTGVNOAOEVIYPT
LKRGKPELVGSESKRGTSPLPAGOVPTLOPOKOVKENTOPPAVADYMPAEL
QYRPPESQYGRPMPRAPYPPPTRRLLPTAPPSRQSKLHEITDKSKBD
TEAWQFPVTLPEMPGEGAOGEPTVEARYKSTIKLXDMKEGVQYGPNSRYMT
LLDSIAHGRLIPYDWEILAKSSLSPOFLQFTKWIIDVOEVRNRANPPVNI
DADLLIGONWSTISQALMONEALEQVRAICLAWEKIDPGSTCPSEFNTVGSXKP
YPPVARLOVAOKSIADIEKARKYIVELAYENANPECOSAIKPLKGVAGSVISE
VYKACDGIIGAMKAMLMADALITGVIVTGRGRCVNCGOIHLKKNCPVLNKN
ITTOATTGRRPPLCPCKKGRKMAQCSKPKRKNOPISGNEORGOPOAPQGTAE
PIQPFVPOGFQGGPPLSVQFGISQLPYNINCPPOAAVQO"
CDS
CDS

/note="translated by frameshift"
/codon_start=3
/product="protease"
/protein_id="AAF88166.1"
/db_xref="GI:9558703"
/translation="NATIVGKAKGASGPTTNWGI PNSAIGSGSGSGPTTPPVSV
GNKPVTTIQQLSPATSGSAVVDCTIQAAVSLGEPQKTPGTVYPLKGVGLILG
RSILNKGQIHITSVSDSDYKGEIDLVSSIPWSAPRDIQAQLLLPLVIGKNSBI
KRIIGIGSTDPGKAAYVASOVSENPVCNAIIQGGQFGLVDYTGADVSIILANQPK
NMPKOKAVGLVIGTASVSEVOSTEILHICLGPNOESVQPMPTSIPLNMGRDLQO
WGAETMPAPSVSPISQKMTMGVIRGGLGNEDGICINPEAKINOREGIGNRC"
CDS
CDS
3878. 6749
/gene="pol"
/gene="pol"
/note="translated by frameshift"
/codon_start=2
/product="polymerase"
/protein_id="AAF88167.1"
/db_xref="GI:9558704"
/translation="NKSRRNRRESILGATVEPPKPIPLTWKTEKPVWVNOPLPQ
KLEALHLANEQLEKHSFSPNSPVFVIOKSGKRMILTDLRANAVIIPMPEL
OGLPSPAMIPKMDWPLIITLDCCPTPLADQCKKFAFTIPAINNKBATPFQKY
LROGMNSTIICOTFVGRALOYRREKSPSCYIHCIDIDILCAETQKIDCTPQOA
EVANAGLAIASDKITQSTPFYHGWQIENRKIKPKQIEIKQDILKTLNDPQKLDGIN
WIRPTGIPTYAMSLFISIRGDSDLNSRMLPEATKIKLVEEKIQSQINRIDL
APLOLIPATYAMSGIILIONDLVMSLPSHVTFTYPLTILQIOLITLQSTRLIK
LGNDDKILVPLETKEOVAFNSGAMKIGLANFVGIIIDNHPKTKI PQLKLTWII
LKPTRREPLNALVTFQSSNGKAYMPKERVYKTPYQASORALVAVITVLD
DOPINISDAYVQATPDRVETALIRYSNDDQNLQNLQOYVRKRNPFYTHIRA
HTNLPGPLKABQADLVLSALIKQELHALHVNAAGKRFVTKWAKOITVQC
TOQOVLHLPTQAGVNRGLCPNALQOMVTHVPSKRFVYAVTVDYSHFIMATQC
TGESSTHVKKHLISCPAMVGVEKIDTDNGPGVCSAPQKFLSQMKSHTGTIPVNSO
GSAIVRTNRILTAKOKEGDSKECTTPQOMLALYTLNPLNINRYOQTTSASQ
HLTKNSPHEGLIMWKNNKTKITIGVITVGRGACVSPENGLPVIPTRHKE
YNEPIADAKSISAEVTSQSTVSDQDNQDNVRRTPDAVHQBRANLGTTHKAD
AVSYKISREHKGITNREYAACSLDDCINGKSPVACRSCS"
CDS
CDS
6451. 8550
/gene="env"
/gene="env"
/codon_start=1
/product="envelope"
/protein_id="AAF88168.1"
/db_xref="GI:9558705"
/translation="NMPSEMKAPPRRRRRHRRARPLTHKNKMTSEBOMKPTSK
ABEPTTAOLKTLTOLATKYLENKVTQTPESMLITVSVVSVLPMPGAAYAT
YNAVYPPPLIAVYTMNDPTEYVYVNDSTWVEPIDDRPAKREBSGMINISIGHY
PPLCLGRAPGCLMPAVQNLVEVPVSPICRFTYHNSGMSLPRVNYLADDSYQSL
KRPKRGKPCPKKIPKESKNTENVLWKECVANSVILLQNEFGTIIIMAPKIGQFVNSG
GOTOSCPASQVSPAVSDLTESLDKHKHKKLQSFYDMEGEEKISTPRPKIVPSGP
EHEMLRLVASHHIRIWSGNTLETFRDKRPFTYIDLSLSITVPLQSCVCPYMLVVG
NIVIKRDSQITICENCRLITCIDSTFNMQHRIILVARGSVWIPVMDRWEASPVH
ILTEVUKGVNRSKRIFITLIVIMELIAYTAAVAGVLAHSVOSVYVNDNMOKRS
TRLNWSQSSIDQKLANQINDLRQTVIMWGBRLMSLHRFQLOCDKMTSDCTIPQYN
ESEHMDWVRHILQGRREDNLTDISLKKQIIFASAAHNLVGTAAIAAGVAGLANL
NPVTWYKTIQSTIILNLIILVCLFLILVCRCTQQLRDRSDHREBAMTMVAVLSRK
GNVGSKSRDQIVTSV"
CDS
CDS
8505. 9472
9475. 9492
/note="lysine primer binding site"
9578. 9586
/note="splice donor site"
9616. 11616
/gene="gag"
9616. 11616
/gene="gag"
/codon_start=1
/product="gag"
/protein_id="AAD21095.2"
/db_xref="GI:9558701"
/translation="MGQTSKISKVASYLUSFIKILKRGVAVSTKNLIKLFQIIEQ
FCMPPEQGLTDLKDKRIKELKQAGKRNIIPLTYNMWAIKALFEPQTEBSV

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CH7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osseogawa, K., Moon, P.Y., Zhao, B., Frangsen, E., Tatenio, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP13-552B11 the clone sequenced to the right is RP11-32P3. 2000 bp overlap. Actual start of this clone is at base position 22189 of RP13-552B11 actual end is at base position 2000 of RP11-32P3.

Location/Qualifiers

1. 154037

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="7"

/map="7"

/clone="RP11-33P21"

/clone_1lb="RPCI-11"

1. 550

/rpt_family="L1"

789. 1317

/rpt_family="ERV.L"

4289. 4432

/rpt_family="MIR"

4964. 5214

/rpt_family="Alu"

5233. 5483

/rpt_family="L1"

5500. 5520

/rpt_family="AT_rich"

5994. 6283

/rpt_family="Alu"

6337. 6661

/rpt_family="Alu"

6991. 7114

/rpt_family="L1"

7401. 7634

/rpt_family="Alu"

7641. 7759

/rpt_family="L2"

7797. 7843

/rpt_family="L1"

7964. 8144

/rpt_family="MaLR"

8148. 8458

/rpt_family="Alu"

8461. 8747

/rpt_family="Alu"

8755. 8806

/rpt_family="(GGGA)n"

9349. 9499

/rpt_family="L1"

9513. 9762

/rpt_family="Alu"

9768. 9936

/rpt_family="Alu"

9938. 10243

repeat_region

/rpt_family="Alu"

repeat_region

10278. 10576

/rpt_family="L1"

repeat_region

10577. 10897

/rpt_family="Alu"

repeat_region

10898. 11020

/rpt_family="L1"

repeat_region

11187. 11411

/rpt_family="L1"

repeat_region

11412. 11711

/rpt_family="Alu"

repeat_region

11712. 11961

/rpt_family="L1"

repeat_region

11973. 12264

/rpt_family="Alu"

repeat_region

12265. 12610

/rpt_family="L1"

repeat_region

12611. 12907

/rpt_family="Alu"

repeat_region

12908. 13739

/rpt_family="L1"

repeat_region

13743. 13876

/rpt_family="L1"

repeat_region

13885. 13934

/rpt_family="MERL_type"

repeat_region

14077. 14139

/rpt_family="MIR"

repeat_region

14576. 14863

/rpt_family="Alu"

repeat_region

15079. 15379

/rpt_family="Alu"

repeat_region

15648. 16009

/rpt_family="Alu"

repeat_region

16015. 16322

/rpt_family="Alu"

repeat_region

16326. 16454

/rpt_family="Alu"

repeat_region

17256. 17384

/rpt_family="L1"

repeat_region

17389. 17628

/rpt_family="Alu"

repeat_region

17639. 18045

/rpt_family="ERV.L"

repeat_region

18046. 18454

/rpt_family="(TA)n"

repeat_region

18455. 18515

/rpt_family="ERV.L"

repeat_region

18516. 18576

/rpt_family="Alu"

repeat_region

18643. 18952

/rpt_family="Alu"

repeat_region

19052. 19447

/rpt_family="MaLR"

Query Match 100.0%; Score 150; DB 9; Length 154037;
Best Local Similarity 100.0%; Pred. No. 1.2e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GAGATAGGGAAAAACCGCCTTAGGCGCTGAGAGTGGGACCTGGGGGAGCAATCTGCTT	60
Db	46857	GAGATAGGGAAAAACCGCCTTAGGCGCTGAGAGTGGGACCTGGGGGAGCAATCTGCTT	46798
Qy	61	GTAAGCAGCTGAGATGTTATGTGTATGATATCTAAAGCAGACGACTTAATCCTTTAC	120
Db	46797	GTAAGCAGCTGAGATGTTATGTGTATGATATCTAAAGCAGACGACTTAATCCTTTAC	46738
Qy	121	ATTGCTATGATGCAAAAGACCTTTGTTAC	150
Db	46737	ATTGCTATGATGCAAAAGACCTTTGTTAC	46708

RESULT 15
AC134684/c

LOCUS AC134684 175202 bp DNA linear PRI 10-APR-2003
 DEFINITION Homo sapiens chromosome 8, clone RP11-1118W6, complete sequence.
 ACCESSION AC134684
 VERSION AC134684.5 GI:29725865
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 175202)
 Birren, B., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 8, clone RP11-1118W6
 Unpublished
 2 (bases 1 to 175202)
 Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
 Karatas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K.,
 Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
 McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Mieng, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
 O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
 Testfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
 JOURNAL Submitted (28-SEP-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 175202)
 Birren, B., Nusbaum, C., Lander, E., Abouneil, A., Allen, N.,
 Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
 Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,
 Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
 Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kelle, C., Landers, T., Levine, R.,
 Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,
 Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
 Meldrim, J., Menus, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J.,
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
 O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
 Talamas, J., Testfaye, S., Theodore, J., Topham, K., Travers, M.,
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
 JOURNAL Submitted (26-MAR-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 175202)
 Birren, B., Nusbaum, C., Lander, E., Abouneil, A., Allen, N.,
 Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
 Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,
 Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
 Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kelle, C., Landers, T., Levine, R.,
 Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,
 Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
 Meldrim, J., Menus, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J.,
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,

TITLE JOURNAL
 COMMENT Submitted (10-APR-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 10, 2003 this sequence version replaced gi:27877285.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WBIR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: l28236
 Center clone name: 1118_M_6

FEATURES
 source
 1. 175202
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="RP11-1118W6"
 /clone_1kb="RP11-11 Human Male BAC"
 544..611
 /rpt_family="CT-rich"
 975..1150
 /rpt_family="AluB"
 2268..2392
 /rpt_family="MIR"
 2633..2798
 /rpt_family="MIR"
 3910..4207
 /rpt_family="AluSx"
 complement(5227..5306)
 /rpt_family="L2"
 6392..6421
 /rpt_family="A)n"
 6842..6897
 /rpt_family="A)n"
 6873
 /rpt_family="TTCC)n"
 /note="Probably C; possibly A"
 complement(7299..7353)
 /rpt_family="MIR3"
 complement(8703..8805)
 /rpt_family="L2"
 9582..9615
 /rpt_family="CAA)n"
 9772..9893
 /rpt_family="MIR"
 9982..10074
 /rpt_family="MIR3"
 complement(10683..10777)
 /rpt_family="L2"
 10778..10932
 /rpt_family="L1M4"
 10999..11082
 /rpt_family="L1M4"
 12172..12205
 /rpt_family="A)n"
 13017..13223
 /rpt_family="L1M5"
 13224..13531
 /rpt_family="AluSg"
 13532..14008

Job time : 629.802 secs

```
repeat_region /rpt_family="L1MB5"
complement(14879..15054)
/rpt_family="L1PA8"
repeat_region /rpt_family="AT_rich"
15180..15206
/rpt_family="MADE1"
repeat_region /rpt_family="MADE1"
16055..16134
/rpt_family="AT_rich"
repeat_region /rpt_family="AT_rich"
16711..16783
17006..17334
/rpt_family="AluSp"
repeat_region /rpt_family="AluSp"
17434..18204
/rpt_family="L1MB3d"
repeat_region /rpt_family="L1MB3d"
18205..18834
/rpt_family="MER51D"
repeat_region /rpt_family="MER51D"
18835..19070
/rpt_family="L1MB3d"
repeat_region /rpt_family="L1MB3d"
complement(24651..24944)
/rpt_family="AluSq"
repeat_region /rpt_family="AluSq"
complement(26607..26682)
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
27183..27520
/rpt_family="L1MC5"
complement(27584..27860)
/rpt_family="L1MC5"
repeat_region /rpt_family="L1MC5"
28775..29075
/rpt_family="L2"
repeat_region /rpt_family="L2"
29234..29336
/rpt_family="L2"
repeat_region /rpt_family="L2"
29353..29466
/rpt_family="L2"
repeat_region /rpt_family="L2"
complement(30426..30581)
/rpt_family="MIR3"
repeat_region /rpt_family="MIR3"
complement(30729..30926)
/rpt_family="MER4B"
repeat_region /rpt_family="MER4B"
complement(30927..31245)
/rpt_family="AluSx"
repeat_region /rpt_family="AluSx"
complement(31246..31293)
/rpt_family="MER4B"
repeat_region /rpt_family="MER4B"
complement(31337..31708)
/rpt_family="MER4B"
repeat_region /rpt_family="MER4B"
32262..32447
/rpt_family="MIR3"
repeat_region /rpt_family="MIR3"
33391..33486
/rpt_family="MIR3"
repeat_region /rpt_family="MIR3"
34024..34052
/rpt_family="AT_rich"
repeat_region /rpt_family="AT_rich"
complement(34310..34400)
/rpt_family="LTR16B"
repeat_region /rpt_family="LTR16B"
34401..34713
/rpt_family="AluDo"
repeat_region /rpt_family="AluDo"
complement(34714..34881)
/rpt_family="LTR16B"
repeat_region /rpt_family="LTR16B"
35338..35931
```

Query Match 100.0%; Score 150; DB 9; Length 175202;
Best Local Similarity 100.0%; Pred.No.1.2e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAGATAGGAGAAACCGCCTTAGGCGCTGAGAGTGGACCTGCGGCGAGCAATACCTGCTT 60
DB 64860 GAGATAGGAGAAACCGCCTTAGGCGCTGAGAGTGGACCTGCGGCGAGCAATACCTGCTT 64801
QY 61 GTAAAGACCTGAGAGTGTATTATGTCATATGATCTAAAGACAGCACTTAATCCTTTAC 120
DB 64800 GTAAAGACCTGAGAGTGTATTATGTCATATGATCTAAAGACAGCACTTAATCCTTTAC 64741
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 64740 ATTGCTATGATGCAAGACCTTTGTTTAC 64711
```

Search completed: February 27, 2004, 23:57:24


```
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Flets, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H. CONTIG"
US-08-724-394A-22

Query Match      81.6%; Score 717.4; DB 2; Length 246240;
Best Local Similarity 91.4%; Pred. No. 5.9e-244;
Matches 805; Conservative 0; Mismatches 71; Indels 5; Gaps 4;

QY 1 TGTGGGAAAAAGCAAGAGATCAGATTGTTCTGTCTGTGTAGAAAGTAGACAT 60
DB 104446 TGTGGGAAAAAGAGAGAGATCAGATTGTTCTGTCTGTGTAGAAAGTAGACAT 104505
QY 61 AGGAGATCCATTTTGTATGACTAGAAATAATCTTCCGCTTGAGATTCTTTATC 120
DB 104506 AGGAGATCCATTTTGTATGACTAGAAATAATCTTCCGCTTGAGATTCTTTATC 104565
QY 121 TATGACCTTACCCCAACCCGCTCTCTGAAACATGCTGTCACCTCAGGGTTAA 180
DB 104566 TA--ACCTTACCCCAACCCGCTCTCTGAAACATGCTGTCACCTCAGGGTTAA 104623
QY 181 ATGATTAAGGGCGGTGCAAGATGCTTTGTTAAACAGATGCTTGAAGCGAGATGCTC 240
DB 104624 ATGATTAAGGGCGGTGCAAGATGCTTTGTTAAAGAAATGCTTGAAGCGAGATGCTC 104683
QY 241 CTTAAGAGTATCAACCACTCCCTTAATCTCAAGTACCAAGGACACAAA--ACTGCCGAG 299
DB 104684 GTTAAAGATATCTTCCATCTCTTAATCTCAAGTACCAAGGACACAAA--ACTGCCGAG 104743
QY 300 GCCGACAGGACCTTCTGCTAGGAAGCAGGATTTGTCCAAGTTTCTCCCATGTGATA 359
DB 104744 GCCGACAGGACCTTCTGCTAGGAAGCAGGATTTGTCCAAGTTTCTCCCATGTGATA 104803
QY 360 GCTGAAATATGCGCTGTGGGAAAGGAAAGACCTGACCGTCCCGACCGGACACCCGT 419
DB 104804 GTTGAATATGCGCTGTGGGAAAGGAAAGACCTGACCGTCCCGACCGGACACCCGT 104863
QY 420 AAAGGCTCTGTGCTGAGAGGATTAAGTAAAGAGAAAGAAATGCTCT--TTGCACTTGA 478
DB 104864 AAAGGCTCTGTGCTGAGAGGATTAAGTAAAGAGAAAGAAATGCTCT--TTGCACTTGA 104923
QY 479 CAAGAGAGAGCATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 538
DB 104924 CAAGAGAGAGCATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 104983
```

```
QY 539 CCGATTGTATGCTCCATCTTACTAGATAGGAAAAAACCCTTAGGCTTGAGAGTGGAC 598
DB 104984 CCGATTGTATTTCCATCTTACTAGATAGGAAAAAACCCTTAGGCTTGAGAGTGGAC 105043
QY 599 CTGCGGAGCAATATCTGCTTTGTAAGACCTGAGATGTTATGTATGATCATATTA 658
DB 105044 ATCTGGACCAATATCTGCTTTGTAAGACCTGAGATGTTATGTATGATCATATCT -AA 105102
QY 659 AGCAGACATTAATCTTACTTACATTTGCTATGATGAGCAAGACCTTTGTCATGTTGT 718
DB 105103 AGCAGACATTAATCTTACTTACATTTGCTATGATGAGCAAGACCTTTGTCATGTTGTAC 105162
QY 719 CTGCTGACCTCTCCCACAATTTGCTTGTGACCCCTGACACATCCCTCTTGAAGAAC 778
DB 105163 CTGCTGACCTCTCTCCCACAATTTGCTTGTGACCCCTGACACATCCCTCTTGAAGAAC 105222
QY 779 ACCACAGATGATCAGTAAATATAAGGAACTCAGAGGCTGCGGATCCTCATATGC 838
DB 105223 ACCACATGATGATCAGTAAATATAAGGAACTCAGAGGCTGCGGATCCTCATATGC 105282
QY 839 TGAACGCTGTTCCCGGGTCCCTTCTTTCTCTATA 879
DB 105283 TGAACACCGGTCCCTGACCCCTTTTCTTTCTCTATA 105323
```

```
RESULT 5
US-09-482-273-20/c
; Sequence 20, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; EARLIER FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 2525
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1354)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-482-273-20
```

```
Query Match      78.7%; Score 692.2; DB 4; Length 2525;
Best Local Similarity 89.8%; Pred. No. 2.4e-236;
Matches 780; Conservative 17; Mismatches 35; Indels 37; Gaps 4;

QY 14 AAGAGATCAGATTGTACTGTGTGTGTAGAAAGAAAGTACATAGAGACTTCATT 73
DB 1725 AAGAGATCAGATTGTACTGTGTGTGTGTAGAAAGAAAGTACATAGAGACTTCATT 1666
QY 74 TTGTTATGATCTAAGAAAAATTTCTTCTGCTTGAGATTCTGTAAATCTATGACTTACC 133
```

```

Db 1665 TTGTTTCG-----AATCTTAACCTTACC 1640
Oy 134 CCAACCCCGTCTCTGAACATGTGCTGTCACCTCAGAGTTAAATGATTAAGGCG 193
Db 1339 CCAACCCCGTCTCTGAACATGTGCTGTCACCTCAGAGTTAAATGATTAAGGCG 1580
Oy 194 GGTGAGAGATGTGTTGTTAAACAGATGCTTGAAGGAGATGCTCTTAAAGTCATC 253
Db 1579 GGTGAGAGATGTGTTGTTAAACAGATGCTTGAAGGAGATGCTCTTAAAGTCATC 1520
Oy 254 ACACCTCCCTTAATCTCAAGTACCCAGGACACAAAA-ACGGGAGAGCCCGAGGACCT 312
Db 1519 ACACCTCCCTTAATCTCAAGTACCCAGGACACAAAA-ACGGGAGAGCCCGAGGACCT 1460
Oy 313 CTGAGTGAAGAGCAGGATTTGTCACAGTTTCCCATGATGATTAAGTAAATATAG 372
Db 1459 CTGAGTGAAGAGCAGGATTTGTCACAGTTTCCCATGATGATGATTAAGTAAATATAG 1400
Oy 373 CCTGCTGGAGAGGAGAGAGCCTGACCTCCCGAGCCGAGACCC-GTAAAGGCTCTGTG 431
Db 1399 CCTGCTGGAGAGGAGAGAGCCTGACCTCCCGAGCCGAGACCCGAGAGGCTCTGTG 1340
Oy 432 CTGAGAGAGATTAGTAAAGAGAGAGAGATGCTCTGAGTGAAGACAGAGAGCA 491
Db 1339 CTGAGAGAGATTAGTAAAGAGAGAGAGATGCTCTGAGTGAAGACAGAGAGCA 1280
Oy 492 TCTGCTCCCTGCTCCCTGCGGAGATGCTGCTGATTAAGAGCCGATTTGATGCT 551
Db 1279 TCTGCTCCCTGCTCCCTGCGGAGATGCTGCTGATTAAGAGCCGATTTGATGCT 1220
Oy 552 CCATCTACTGAGATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 611
Db 1219 CCATCTACTGAGATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1160
Oy 612 TACTGCTTTGTAAGACAGTGAATGTTATGATGATGATGATGATGATGATGATGAT 671
Db 1159 TACTGCTTTGTAAGACAGTGAATGTTATGATGATGATGATGATGATGATGATGAT 1100
Oy 672 ATCTTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 731
Db 1099 ATCTTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1040
Oy 732 CCCCACATTTGCTTTGACCTGACATCCCTCTTCCAGAGAGAGAGAGAGAGAGAGAT 791
Db 1039 CCCCACATTTGCTTTGACCTGACATCCCTCTTCCAGAGAGAGAGAGAGAGAGAT 980
Oy 792 CAGTAAATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 850
Db 979 CAGTAAATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 920
Oy 851 CCCCCGGGCTCTTTCTTTCTTTCTTATA 879
Db 919 CCCCCGGGCTCTTTCTTTCTTTCTTATA 891

```

RESULT 6
 US-08-991-789A-141
 ; Sequence 141, Application US/08991789A
 ; Patent No. 6225054
 ; GENERAL INFORMATION:
 ; APPLICANT: Prudkier, Tony N.
 ; ; Smith, John M.
 ; ; Reed, Steven G.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 ; TREATMENT AND DIAGNOSIS OF BREAST CANCER
 ; NUMBER OF SEQUENCES: 292
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed IP Law Group
 ; STREET: 701 Fifth Avenue, Suite 6300
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/991,789A
 FILING DATE: 11-Dec-1997
 CLASSIFICATION: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Potter, Jane E. R.
 REGISTRATION NUMBER: 33,332
 REFERENCE/DOCKET NUMBER: 210121.419C3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 141:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9388 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 141:
 US-08-991-789A-141

Query Match 56.3%; Score 494.8; DB 3; Length 9388;
 Best Local Similarity 79.1%; Pred. No. 2.1e-165;
 Matches 709; Conservative 0; Mismatches 132; Indels 55; Gaps 8;

```

Oy 20 GATCAGATTGTACTGTGTCTGTGTAGAGAGAGATGATGATGATGATGATGATGATGAT 79
Db 48 GATCAGATTGTACTGTGTCTGTGTAGAGAGAGATGATGATGATGATGATGATGATGAT 107
Oy 80 TGTACTAAGAAAAATTTCTTCTGCTTGAATTTCTTATGATGATGATGATGATGATGAT 139
Db 108 TGTACTAAGAAAAATTTCTTCTGCTTGAATTTCTTATGATGATGATGATGATGATGAT 167
Oy 140 CCGGCTCTCTGAAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 199
Db 168 CTGCTCTCAGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 221
Oy 200 GATGCTGCTTGTAAACAGATGCTTGAAGGAGAGATGCTTGAAGGAGATGCTTGAAGG 259
Db 222 ---TATGCTTTGTTAAAGAGATGCTTGAAGGAGATGCTTGAAGGAGATGCTTGAAGG 278
Oy 260 CCTTAATCTCAAGTACCCAGGAGACAA-AACTGCGAAGGCGCAGGAGCTTCTGCT 318
Db 279 CTTAATCTCAAGTACCCAGGAGACAAATACTGCGAAGGCGCAGGAGCTTCTGCT 338
Oy 319 AGGAAAGCAGGATTTGTCACAGCTTCTCCCAATGATGATGATGATGATGATGATGATG 378
Db 339 AGGAAAGCAGGATTTGTCACAGCTTCTCCCAATGATGATGATGATGATGATGATGATG 398
Oy 379 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 408
Db 399 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 458
Oy 409 CCGACACCCGTAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 468
Db 459 CCGACACCCGTAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 515
Oy 469 GCAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 528
Db 516 GCAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 575
Oy 529 GGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 588
Db 576 GGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 635
Oy 589 GAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 648
Db 636 GAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 695

```


SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/598,326
FILING DATE: 20-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E.R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 141:
SEQUENCE CHARACTERISTICS:
LENGTH: 9388 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 141:
US-09-598-326-141

Query Match 56.3%; Score 494.8; DB 4; Length 9388;
Best Local Similarity 79.1%; Pred. No. 2.1e-165;
Matches 709; Conservative 0; Mismatches 132; Indels 55; Gaps 8;

Qy 20 GATCAGATTGTTACTGTCGTCTGTGAGAAAGATAGACATAGAGACTCCATTTTGTGA 79
Db 48 GATAGACTGTTACTGTCGTCTATATGAAAGATAGACATAGAGATTCATTTTGTTC 107
Qy 80 TGTACTAAGAAAAATTTCTTGCCTTGATTTCTTATCTATGACCTTACCCCAACC 139
Db 108 TGTACTAAGAAAAATTTCTTGCCTTGATTTCTTATCTGTAACCTTACCCCAACC 167
Qy 140 CCGTCTCTCTGAAACATGTCGTCTGTCCACTCAGGGTTAAATGATTAAAGGGGGTGA 199
Db 168 CTGTGCTCACAAGACATGTCGTCTGTGACTCAAGGTTCAATGATTAGGGC----- 221
Qy 200 GGATGTCCTTTGTTAAACAGATGCTGGAAGGACAGATGCTTAAAGATGATCAACACT 259
Db 222 ---TATGCTTTGTTAAAAAGTCTTGAAGATTAATGCTTTAAAGTATCAACACT 278
Qy 260 CCTTAATCTCAAGTACCAGGAGACAA-AAAATGCGGAAGCGGAGGAGACTCTGCT 318
Db 279 CTCTAATCTCAAGTACCAGGAGACAAATACATGCGGAAGCGGAGGAGACTCTGCT 338
Qy 319 AGGAAAGCCAGTATTGTCCAGCTTCTCCCATGTGATAGCTTGAATATGACCTGCT 378
Db 339 AGGAAAGCCAGTATTGTCCAGATTTCTCCCATGTGATAGCTTGAATATGACCTGCT 398
Qy 379 GGGAAAGGAAAGACTGAC-----CGTCCCAAGC 408
Db 399 GGGAAAGGTAAGACTGCTGTCCTCCCAAGCCGACATCCCAAGCCGACATCCCAAGC 458
Qy 409 CCGACACCCGTAAGAGGCTGTGCTGAGAGATTAAGTAAAGAGAAAGATCTCTT 468
Db 459 CCGACACCCGTAAGAGGCTGTGCTGAGAGATTAAGTAAAGAGAAAG---CCTCTT 515
Qy 469 GCACTTGAACAGAGAGAGGCACTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 528
Db 516 GCACTTGAAGTAAGAGAGAGGCACTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 575
Qy 529 GGTATAAAACCCGTTGATGCTCCATCTACTGAGATAGGAAAAACCGCTTGGGCTG 588
Db 576 GGTATAAAACCCGTTGATGCTTACTTACTGAGATAGGAAAAACCTTTAGGGCTG 635
Qy 589 GAGTGGGACCTGCGGAGACATATGCTTTGTAAGCACTGAGATGTTATGTATG 648
Db 636 GAGTGGAGACAGCTGCGGAGATATGCTTTTAAAGCAAGATGTTTGTATAGT 695
Qy 649 CATATCTAAAGACAGACTTAAATCTTTAATTTGCTATGATGACAAAGACTTTGTT 708
Db 696 CACATC--AAGGACAGACACT--TTCCCTAACTTATATGACAGAGACCTTTGTT 752

Qy 709 ACATGTTGTCGTGACCCCTCTCCCAACAATGCTTTGAGACCTTGACACATCCCTTC 768
Db 753 AC--GTTTTCGTGACCCCTCTCCCACTATATACCATTTAGGCTGACATCCCTTC 810
Qy 769 TTCGAAACACCCAGATGATGATTAATATCTAAGGAATCTAGAGCTGGG----- 823
Db 811 TCCGAGATGTAGAGATTAATGATCAATAAATCTAGAGGAATCTAGAGAGATGCTCCT 870
Qy 824 --GGATCCTCATATGCTGAAAGCTGATCCCGGATCCCTCTTTCTTTCTCTGA 877
Db 871 GTAGTCTCTCGTGTGCTGACGCGGATCTTGGGCTCACTTTCTTTCTATTA 926

RESULT 9
US-09-289-198-141
Sequence 141: Application US/09289198
Patent No. 6586570
GENERAL INFORMATION:
APPLICANT: Frudakis, Tony N.
APPLICANT: Smith, John M.
APPLICANT: Reed, Steven G.
APPLICANT: Misher, Lynda
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TITILE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER
FILE REFERENCE: 210121.419C5
CURRENT APPLICATION NUMBER: US/09/289,198
CURRENT FILING DATE: 1999-04-09
EARLIER APPLICATION NUMBER: US 09/062,451
EARLIER FILING DATE: 1998-04-17
EARLIER APPLICATION NUMBER: US 08/991,789
EARLIER FILING DATE: 1997-12-11
EARLIER APPLICATION NUMBER: US 08/838,762
EARLIER FILING DATE: 1997-04-09
EARLIER APPLICATION NUMBER: PCT/US97/00485
EARLIER FILING DATE: 1997-01-10
EARLIER APPLICATION NUMBER: US 08/700,014
EARLIER FILING DATE: 1996-08-20
EARLIER APPLICATION NUMBER: US 08/585,392
EARLIER FILING DATE: 1996-01-01
NUMBER OF SEQ ID NOS: 312
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 141
LENGTH: 9388
TYPE: DNA
ORGANISM: Homo sapien
US-09-289-198-141

Query Match 56.3%; Score 494.8; DB 4; Length 9388;
Best Local Similarity 79.1%; Pred. No. 2.1e-165;
Matches 709; Conservative 0; Mismatches 132; Indels 55; Gaps 8;

Qy 20 GATCAGATTGTTACTGTCGTCTGTGAGAAAGATAGACATAGAGACTCCATTTTGTGA 79
Db 48 GATCAGACTGTTACTGTCGTCTATATGAAAGATAGACATAGAGATTCATTTTGTTC 107
Qy 80 TGTACTAAGAAAAATTTCTTGCCTTGATTTCTTATCTATGACCTTACCCCAACC 139
Db 108 TGTACTAAGAAAAATTTCTTGCCTTGATTTCTTATCTGTAACCTTACCCCAACC 167
Qy 140 CCGTCTCTCTGAAACATGTCGTCTGTCCACTCAGGGTTAAATGATTAAAGGGGGTGA 199
Db 168 CTGTGCTCACAAGACATGTCGTCTGTGACTCAAGGTTCAATGATTAGGGC----- 221
Qy 200 GGATGTCCTTTGTTAAACAGATGCTGGAAGGACAGATGCTTAAAGATGATCAACACT 259
Db 222 ---TATGCTTTGTTAAAAAGTCTTGAAGATTAATGCTTTAAAGTATCAACACT 278
Qy 260 CCTTAATCTCAAGTACCAGGAGACAA-AAAATGCGGAAGCGGAGGAGACTCTGCT 318
Db 279 CTCTAATCTCAAGTACCAGGAGACAAATACATGCGGAAGCGGAGGAGACTCTGCT 338
Qy 319 AGGAAAGCCAGTATTGTCCAGCTTCTCCCATGTGATAGCTTGAATATGACCTGCT 378

FEATURE:
NAME/KEY: CDS
LOCATION: 119..334
US-09-621-976-2586

Query Match 51.0%; Score 448.6; DB 4; Length 482;
Best Local Similarity 96.9%; Pred. No. 8.3e-150;
Matches 468; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

```

QY 84 CTAAGAAAAATTTCTTCCCTTGAATTCCTGATCTATATACCTTACCCCAACCCCGT 143
DB 1 CTAAGAAAAATTTCTTCCCTTGAATTCCTGATCTATATACCTTACCCCAACCCCGT 60
QY 144 GCTCTGAAACATGTGCTGTGCTCACTCAGAGTTAAATGATTAAAGGCGGTGAGAGT 203
DB 61 GCTCTGAAACATGTGCTGTGCTCACTCAGAGTTAAATGATTAAAGGCGGTGAGAGT 120
QY 204 GTGCTTTGTTAAAGATGCTTGAAGGAGCATGCTCTTAAGATCACTCACTCCCT 263
DB 121 GTGCTTTGTTAAAGATGCTTGAAGGAGCATGCTCTTAAGATCACTCACTCCCT 180
QY 264 AATCTCAAGTACCAAGGACACAAAATGCGGAAGGCGGAGGAGCTTGTCTAGGAA 323
DB 181 AATCTCAAGTACCAAGGACACAAAATGCGGAAGGCGGAGGAGCTTGTCTAGGAA 239
QY 324 AGCAGGATTTGTCACAGTTTCTCCCATGTGATAGCTGAAATATAGGCTCGTGGAA 383
DB 240 AGCAGGATTTGTCACAGTTTCTCCCATGTGATAGCTGAAATATAGGCTCGTGGAA 299
QY 384 GGAAGAACTGACCGTCCCGGACCCGACACCGGTAAGGCTGTGCTGAGAGATT 443
DB 300 GGAAGAACTGACCGTCCCGGACCCGACACCGGTAAGGCTGTGCTGAGAGATT 359
QY 444 AGTAAAGAGAGAAATGCTCTTTCAGTTGAGACAAAGAGGATCTGTCTCTGC 503
DB 360 AGTAAAGAGAGAAATGCTCTTTCAGTTGAGACAAAGAGGATCTGTCTCTGC 419
QY 504 CTGCTCCGCGCAATGGAATGCTCGGATTAATAAACCGGATTTGATGCTCACTAGAG 563
DB 420 CTGCTCCGCGCAATGGAATGCTCGGATTAATAAACCGGATTTGATGCTCACTAGAG 479
QY 564 ATA 566
DB 480 ATA 482

```

RESULT 12
US-09-621-976-2859
Sequence 2859, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET 054PR2
CURRENT APPLICATION NUMBER: US/09/621.976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 2859
LENGTH: 488
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 119..334
US-09-621-976-2859

Query Match 50.4%; Score 442.8; DB 4; Length 488;
Best Local Similarity 96.5%; Pred. No. 9.9e-148;
Matches 473; Conservative 1; Mismatches 14; Indels 2; Gaps 2;

```

QY 84 CTAAGAAAAATTTCTTCCCTTGAATTCCTGATCTATATACCTTACCCCAACCCCGT 143
DB 1 CTAAGAAAAATTTCTTCCCTTGAATTCCTGATCTATATACCTTACCCCAACCCCGT 60
QY 144 GCTCTGAAACATGTGCTGTGCTCACTCAGAGTTAAATGATTAAAGGCGGTGAGAGT 203
DB 61 GCTCTGAAACATGTGCTGTGCTCACTCAGAGTTAAATGATTAAAGGCGGTGAGAGT 120
QY 204 GTGCTTTGTTAAAGATGCTTGAAGGAGCATGCTCTTAAGATCACTCACTCCCT 263
DB 121 GTGCTTTGTTAAAGATGCTTGAAGGAGCATGCTCTTAAGATCACTCACTCCCT 180
QY 264 AATCTCAAGTACCAAGGACACAAAATGCGGAAGGCGGAGGAGCTTGTCTAGGAA 323
DB 181 AATCTCAAGTACCAAGGACACAAAATGCGGAAGGCGGAGGAGCTTGTCTAGGAA 239
QY 324 AGCAGGATTTGTCACAGTTTCTCCCATGTGATAGCTGAAATATAGGCTCGTGGAA 383
DB 240 AGCAGGATTTGTCACAGTTTCTCCCATGTGATAGCTGAAATATAGGCTCGTGGAA 299
QY 384 GGAAGAACTGACCGTCCCGGACCCGACACCGGTAAGGCTGTGCTGAGAGATT 443
DB 300 GGAAGAACTGACCGTCCCGGACCCGACACCGGTAAGGCTGTGCTGAGAGATT 359
QY 444 AGTAAAGAGAGAAATGCTCTTTCAGTTGAGACAAAGAGGATCTGTCTCTGC 503
DB 360 AGTAAAGAGAGAAATGCTCTTTCAGTTGAGACAAAGAGGATCTGTCTCTGC 419
QY 504 CTGCTCCGCGCAATGGAATGCTCGGATTAATAAACCGGATTTGATGCTCACTAGAG 563
DB 420 CTGCTCCGCGCAATGGAATGCTCGGATTAATAAACCGGATTTGATGCTCACTAGAG 478
QY 564 ATAGGAAA 573
DB 479 ATAGGAAA 488

```

RESULT 13
US-09-328-111-600
Sequence 600, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Steimann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Derti, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328.111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 600
LENGTH: 589
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(589)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-600

Query Match 41.2%; Score 362; DB 3; Length 589;
Best Local Similarity 78.6%; Pred. No. 8,4e-119;
Matches 469; Conservative 0; Mismatches 120; Indels 8; Gaps 3;

QY 274 ACCGAGGACACAAAACCTGGAGAGCCGAGGAGCTGCTAGAGAAAGCAGATAT 333
DB 1 ACCGAGGACACAAAACCTGGAGAGCTGAGAGGAGCTGCTAGAGAAAGCAGATAT 60
QY 334 TGTCCACGTTTCTCCCATGTGATAGCTGAAATATGAGCTGAGGAGGAGAAAGACC 393
DB 61 TGTCCACGTTTCTCCCATGTGATAGCTGAAATATGAGCTGAGGAGGAGAAAGACC 120
QY 394 TGACCGTCCCGCAGCCGACACCCGTAAGAGGCTGTGCTGAGAGAGATTAGTAAAGAG 453
DB 121 TGACCGTCCCGCAGCCGACACCCATTAAGAGGCTGTGCTGAGAGAGATTAGTAAAGAG 180
QY 454 GAAGGATGCTCTTGATGATGAGACAAAGAGAGAGATCTGTCTCTGCTGCTGCTG 513
DB 181 GAAGGATGCTCTTGATGATGAGATTAAGAGAGAGATCTGTCTCTGCTGCTGCTG 237
QY 514 GCATGAGATGCTCTGATTAACCCGATTTGATGCTCATCTAGATAGAGAGAA 573
DB 238 GCATGAGATGCTCTGATTTAAACCCGATTTGATGCTCATCTAGATAGAGAGAA 297
QY 574 ACCGCTTAAAGGCTGAGAGTGGAGCTGCGGAGACAAATCTGTTTAAAGCACTGAG 633
DB 298 ACTGCTTAAAGGCTGAGAGTGGAGATGCTGCTGAGCAATCTGCTCTTAATGATGAG 357
QY 634 ATGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 693
DB 358 ATGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 415
QY 694 CAAAGACCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 753
DB 416 CAAAGACCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 475
QY 754 TGACACATCCCTCTGAGAGAAACCCACAGATGATGATGATGATGATGATGATG 813
DB 476 CCACATCCCTCTGAGAGAAACCCACAGATGATGATGATGATGATGATGATGATG 532
QY 814 GAGGCTGCGGAGATCTGATGATGATGATGATGATGATGATGATGATGATGATG 870
DB 533 ANAGGCTGCGGAGATCTGATGATGATGATGATGATGATGATGATGATGATGATG 589

RESULT 14
US-08-991-789A-7
Sequence 7, Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:
APPLICANT: Firdakis, Tony N.
Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:

NAME: Potter, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1086 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-991-789A-7

Query Match 40.5%; Score 355.6; DB 3; Length 1086;
Best Local Similarity 80.4%; Pred. No. 2.5e-116;
Matches 532; Conservative 0; Mismatches 81; Indels 49; Gaps 8;

QY 20 GATCAGATTGTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 79
DB 54 GATCAGATTGTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 113
QY 80 TGTACTAAGAAAAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 139
DB 114 TGTACTAAGAAAAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 173
QY 140 CCGTGTCTCTGTAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 199
DB 174 CCGTGTCTCTGTAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 227
QY 200 GATGCTGCTTGTAAACAGATGCTTGAAGAGCAGATGCTCTTAAAGTCACTACCACT 259
DB 228 ---TATGCTTGTAAACAGATGCTTGAAGAGCAGATGCTCTTAAAGTCACTACCACT 284
QY 260 CCTAATCTCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 318
DB 285 CTTAATCTCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 344
QY 319 AGGAAAGCCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 378
DB 345 AGGAAAGCCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 404
QY 379 GGGAGGAGAAAGCCTGACCG-----TCCCCAGC 408
DB 405 GGGAGGAGGTAAGCCTGACCTGCTCCCGACCGACATCCCCAGCCGACATCCCCAGC 464
QY 409 CCGACACCCGTAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 468
DB 465 CCGACACCCGTAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 519
QY 469 GCAATTGAGACAGAGAGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 528
DB 520 TGCATTGAGTAAGAGAGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 578
QY 529 GGTATTAACCCGATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 587
DB 579 GGTATTAACCCGATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 638
QY 588 GAGAGTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 647
DB 639 GAGAGTGGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 696
QY 648 GC 649
DB 697 GC 698

RESULT 15
US-09-062-451-7
Sequence 7, Application US/09062451
Patent No. 6344550
GENERAL INFORMATION:

this page blank (us010)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2004, 21:22:12 ; Search time 77.9883 Seconds
(without alignments)
8170.833 Million cell updates/sec

Title: US-10-016-604-155

Perfect score: 150

Sequence: 1 gagataggggaaaaacgcctc.....atgcaagacccttcttcac 150

Scoring table: IDENTITY NUC

Gap 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

1: geneseqn1980s.*
2: geneseqn1990s.*
3: geneseqn2000s.*
4: geneseqn2001as.*
5: geneseqn2001bs.*
6: geneseqn2002s.*
7: geneseqn2003as.*
8: geneseqn2003bs.*
9: geneseqn2003cs.*
10: geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	100.0	150	6	ABX04696 Human end
2	150	100.0	150	9	ADCl5773 Putative
3	150	100.0	319	6	ABX04570 Human end
4	150	100.0	319	9	ADCl6256 Human her
5	150	100.0	408	6	ABX04571 Human her
6	150	100.0	408	9	ADCl6257 Human her
7	150	100.0	879	6	ABX04572 Human end
8	150	100.0	879	9	ADCl6258 Human her
9	150	100.0	968	6	ABX04691 Human end
10	150	100.0	968	9	ADCl5768 Human end
11	148.4	98.9	2431	4	AAK94873 Human ful
12	148.4	98.9	6556	4	AAK70301 Human imm
13	148.4	98.9	35414	3	AAAD00147 Human reia
14	148.4	98.9	222930	6	ABK84349 Human end
15	146.8	97.9	968	6	ABX04695 Human end
16	146.8	97.9	968	9	ADCl5771 Human end
17	146.8	97.9	968	6	ABX04694 Human end
18	146.8	97.9	968	9	ADCl5772 Human end
19	146.8	97.9	968	9	ADCl5770 Human end
20	146.8	97.9	968	9	ADCl5771 Human end
21	146.8	97.9	32249	4	AAI05336 Human rep
22	146.8	97.9	32249	4	ABL98205 Human tes
23	146	97.3	2525	3	AAZ98026 Human sec

C	24	146	97.3	2525	4	AAD11639 Human sec
C	25	146	97.3	2525	6	ABK69735 Human sec
C	26	145.2	96.8	1659	5	AA666662 DNA encod
C	27	145.2	96.8	1701	5	AA833303 DNA encod
C	28	145.2	96.8	1704	5	AA833549 DNA encod
C	29	145.2	96.8	1782	5	AA879381 DNA encod
C	30	145.2	96.8	1826	5	AA870436 DNA encod
C	31	145.2	96.8	3658	5	AA870654 DNA encod
C	32	145.2	96.8	9179	6	ABX04612 Human pro
C	33	145.2	96.8	9659	6	AA157703 Human pro
C	34	145.2	96.8	9659	8	ADA02672 Human PRL
C	35	145.2	96.8	9659	9	ADB72410 Human PRL
C	36	145.2	96.8	154902	6	ABQ88198 Human ost
C	37	143.6	95.7	150	9	ADCl5738 TAR nucle
C	38	143.6	95.7	962	6	ABX04692 Human end
C	39	143.6	95.7	962	9	ADCl5769 HML-2 LTR
C	40	143.6	95.7	139904	6	ABK83562 Human .CDN
C	41	143.6	95.7	144792	9	ADC87620 Human GPC
C	42	143.6	95.7	349981	9	ADC87619 Human GPC
C	43	140.4	93.6	975	7	ACC43227 Nucleotid
C	44	140.4	93.6	2689	7	ACC43223 Nucleotid
C	45	140.4	93.6	6020	7	ACC43236 Nucleotid

ALIGNMENTS

RESULT 1	ABX04696 standard; DNA, 150 BP.
AC	ABX04696;
XX	
DT	14-JAN-2003 (first entry)
XX	
DE	Human endogenous retrovirus k (herv-k) 5' RUS region.
XX	
KW	Human; endogenous retrovirus; herv; prostate cancer; testicular cancer; multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease; cancer; transgenic animal; ds.
KM	
OS	Human endogenous retrovirus.
XX	
PN	WO200246477-A2.
XX	
PD	13-JUN-2002.
XX	
PF	07-DEC-2001; 2001WO-US047824.
XX	
PR	07-DEC-2000; 2000US-0251830P.
XX	
PR	07-DEC-2001; 2001US-00016604.
XX	
PA	(CHIR) CHIRON CORP.
XX	
PI	Garcia P, Hardy SF, Williams LT, Escobedo J;
XX	
DR	WPI; 2002-691475/74.
XX	
PT	Novel isolated polypeptides useful for diagnosis of prostate cancer.
XX	
PS	Claim 4; Page 149; 152pp; English.
XX	
CC	The invention describes novel isolated polypeptides (I, Ib) useful for
CC	diagnosing prostate cancer comprising obtaining a patient sample
CC	containing prostate cells and detecting the presence or absence of an
CC	expression product of a HML-2 endogenous retrovirus in a patient sample.
CC	Polynucleotides associated with (I) are useful for diagnosis or treatment
CC	of testicular cancer, multiple sclerosis or insulin-dependent diabetes
CC	mellitus. An inhibitor of a HML-2 protease and a transdominant negative
CC	mutant of HML-2 GORP are also useful in the manufacture of a medicament
CC	for treating prostate cancer. (I) and (Ib) are useful for generating
CC	antibodies specific to the polypeptides associated with cancer, as
CC	targets for therapeutic intervention, and in immunising a transgenic

CC animal. This sequence represents a region of human endogenous retrovirus
CC (herv) of the HML-2 sub-group
XX
SQ Sequence 150 BP; 43 A; 29 C; 37 G; 41 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 6; Length 150;
Best Local Similarity 100.0%; Pred. No. 4.1e-41;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGGAACCCGCTTAGGGCTGAGGAGGAGGACCTGCGGGAGCAATCTGCTTT 60
DB 1 GAGATAGGGAACCCGCTTAGGGCTGAGGAGGAGGACCTGCGGGAGCAATCTGCTTT 60
QY 61 GTAAGGACCTGAGATGTTTATGTATGATGATCTAAAGCAGACACTTAATCCTTTAC 120
DB 61 GTAAGGACCTGAGATGTTTATGTATGATGATCTAAAGCAGACACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAAAGACCTTTGTTTAC 150
DB 121 ATTGCTATGATGCAAAAGACCTTTGTTTAC 150

RESULT 2
ADCI5773
ID ADCI5773 standard; DNA; 150 BP.

XX AC ADCI5773;

DT 18-DEC-2003 (first entry)

DE Putative TAR of herv-k(hml-2.hom) nucleotide sequence SEQ ID NO:49.

XX KM prostate cancer; HML-2; env; cytosolic; neuroprotective; antidiabetic;
XX KM immunostimulant; vaccine; gene therapy; breast cancer; testicular cancer;
XX KM multiple sclerosis; insulin-dependent diabetes mellitus; human; gene; ds.

XX OS Homo sapiens.

XX PN WO2003050258-A2.

XX PD 19-JUN-2003.

XX PF 09-DEC-2002; 2002WO-US039344.

XX PR 07-DEC-2001; 2001US-00016604.

XX PR 07-DEC-2001; 2001US-0340664P.

XX PR 12-JUN-2002; 2002US-0388046P.

XX PA (CHIR) CHIRON CORP.

XX PI Garcia P, Hardy SF, Williams LT, Escobedo J;

XX DR WPI; 2003-577296/54.

XX PT Diagnosing prostate cancer, comprises detecting the presence or absence
XX of HML-2 (human endogenous retrovirus) expression product in a patient
XX sample.

XX PS Disclosure; SEQ ID NO 49; 117bp; English.

XX CC The present invention describes a method for diagnosing prostate cancer,
XX comprising detecting the presence or absence of HML-2 expression product
XX in a patient sample, where the expression product is produced by a
XX splicing event in which the 5' region and start codon of the env coding
XX region are joined to a downstream coding region in the reading frame +2
XX relative to that of env. Also described: (1) an isolated polynucleotide;
XX (2) an isolated polypeptide; (3) an antibody that binds to the
XX polypeptide; and (4) an immunogenic composition. Prostate cancer related
XX sequences have cytosolic, neuroprotective, antidiabetic and
XX immunostimulant activities, and can be used in vaccines and in gene
XX therapy. A polynucleotide, polypeptide or antibody of the invention can
XX be used in the manufacture of a medicament for preventing or treating

CC prostate, breast or testicular cancer, multiple sclerosis or insulin-
CC dependent diabetes mellitus. The present sequence represents a sequence
CC which is used in the exemplification of the present invention.
XX
SQ Sequence 150 BP; 43 A; 29 C; 37 G; 41 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 9; Length 150;
Best Local Similarity 100.0%; Pred. No. 4.1e-41;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGGAACCCGCTTAGGGCTGAGGAGGAGGACCTGCGGGAGCAATCTGCTTT 60
DB 1 GAGATAGGGAACCCGCTTAGGGCTGAGGAGGAGGACCTGCGGGAGCAATCTGCTTT 60
QY 61 GTAAGGACCTGAGATGTTTATGTATGATGATCTAAAGCAGACACTTAATCCTTTAC 120
DB 61 GTAAGGACCTGAGATGTTTATGTATGATGATCTAAAGCAGACACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAAAGACCTTTGTTTAC 150
DB 121 ATTGCTATGATGCAAAAGACCTTTGTTTAC 150

RESULT 3
ABX04570
ID ABX04570 standard; DNA; 319 BP.

XX AC ABX04570;

DT 14-JAN-2003 (first entry)

DE Human endogenous retrovirus k (herv-k) R region.

XX KM Human endogenous retrovirus; herv; prostate cancer; testicular cancer;
XX KM multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
XX KM cancer; transgenic animal; ds.

XX OS Human endogenous retrovirus.

XX PN WO200246477-A2.

XX PD 13-JUN-2002.

XX PF 07-DEC-2001; 2001WO-US047824.

XX PR 07-DEC-2000; 2000US-0251830P.

XX PR 07-DEC-2001; 2001US-00016604.

XX PA (CHIR) CHIRON CORP.

XX PI Garcia P, Hardy SF, Williams LT, Escobedo J;

XX DR WPI; 2002-691475/74.

XX PT Novel isolated polypeptides useful for diagnosis of prostate cancer.
XX Disclosure; Page 129; 152pp; English.

XX CC The invention describes novel isolated polypeptides (I, II) useful for
XX diagnosing prostate cancer comprising obtaining a patient sample
XX containing prostate cells and detecting the presence or absence of an
XX expression product of a HML-2 endogenous retrovirus in a patient sample.
XX Polynucleotides associated with (I) are useful for diagnosis or treatment
XX of testicular cancer, multiple sclerosis or insulin-dependent diabetes
XX mellitus. An inhibitor of a HML-2 protease and a transdominant negative
XX mutant of HML-2 cORF are also useful in the manufacture of a medicament
XX for treating prostate cancer. (I) and (II) are useful for generating
XX antibodies specific to the polypeptides associated with cancer, as
XX targets for therapeutic intervention, and in immunising a transgenic
XX animal. This sequence represents a region of human endogenous retrovirus
XX (herv) of the HML-2 sub-group
XX
SQ Sequence 319 BP; 79 A; 83 C; 69 G; 88 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 6; Length 319;
 Best Local Similarity 100.0%; Pred. No. 5.4e-41;
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGGAAAAACCCCTTTAGGGCTGAGGTGGACCTTGGCGGACGCAATATCTGCTTT 60
 DB 1 GAGATAGGGAAAAACCCCTTTAGGGCTGAGGTGGACCTTGGCGGACGCAATATCTGCTTT 60
 QY 61 GTAAAGCAGCTGAGATGTTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
 DB 61 GTAAAGCAGCTGAGATGTTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
 QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
 DB 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150

RESULT 4
 ADC16256
 ID ADC16256 standard; DNA; 319 BP.

AC ADC16256;
 DT 18-DEC-2003 (first entry)
 DE Human herv-k(hml-2.hom) R region nucleotide sequence SEQ ID NO:3.

KW prostate cancer; HML-2; env; cytostatic; neuroprotective; antidiabetic;
 KW immunostimulant; vaccine; gene therapy; breast cancer; testicular cancer;
 KW multiple sclerosis; insulin-dependent diabetes mellitus; human; gene; ds.
 OS Homo sapiens.

PN WO2003050258-A2.

PD 19-JUN-2003.

PF 09-DEC-2002; 2002WO-US039344.

PR 07-DEC-2001; 2001US-00016604.

PR 07-DEC-2001; 2001US-034064P.

PR 07-DEC-2001; 2001WO-US047824.

PR 12-JUN-2002; 2002US-0388046P.

PA (CHIR) CHIRON CORP.

PI Garcia P, Hardy SF, Williams LT, Escobedo J;

DR WPI; 2003-577296/54.

PT Diagnosing prostate cancer, comprises detecting the presence or absence

PT of HML-2 (human endogenous retrovirus) expression product in a patient

PT sample.

PS Disclosure; SEQ ID NO 3; 117pp; English.

CC The present invention describes a method for diagnosing prostate cancer,
 CC comprising detecting the presence or absence of HML-2 expression product
 CC in a patient sample, where the expression product is produced by a
 CC splicing event in which the 5' region and start codon of the env coding
 CC region are joined to a downstream coding region in the reading frame +2
 CC relative to that of env. Also described: (1) an isolated polynucleotide;
 CC (2) an isolated polypeptide; (3) an antibody that binds to the
 CC polypeptide; and (4) an immunogenic composition. Prostate cancer related
 CC sequences have cytostatic, neuroprotective, antidiabetic and
 CC immunostimulant activities, and can be used in vaccines and in gene
 CC therapy. A polynucleotide, polypeptide or antibody of the invention can
 CC be used in the manufacture of a medicament for preventing or treating
 CC prostate, breast or testicular cancer, multiple sclerosis or insulin-
 CC dependent diabetes mellitus. The present sequence represents a sequence
 CC which is used in the exemplification of the present invention.

SO Sequence 319 BP; 79 A; 83 C; 69 G; 88 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 9; Length 319;
 Best Local Similarity 100.0%; Pred. No. 5.4e-41;
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGGAAAAACCCCTTTAGGGCTGAGGTGGACCTTGGCGGACGCAATATCTGCTTT 60
 DB 1 GAGATAGGGAAAAACCCCTTTAGGGCTGAGGTGGACCTTGGCGGACGCAATATCTGCTTT 60
 QY 61 GTAAAGCAGCTGAGATGTTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
 DB 61 GTAAAGCAGCTGAGATGTTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
 QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
 DB 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150

RESULT 5
 ABX04571
 ID ABX04571 standard; DNA; 408 BP.

AC ABX04571;
 DT 14-JAN-2003 (first entry)

DE Human endogenous retrovirus k (herv-k) RUS region.

KW Human; endogenous retrovirus; herv; prostate cancer; testicular cancer;
 KW multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
 KW cancer; transgenic animal; ds.

OS Human endogenous retrovirus.

PN WO200246477-A2.

PD 13-JUN-2002.

PF 07-DEC-2001; 2001WO-US047824.

PR 07-DEC-2000; 2000US-0251830P.

PR 07-DEC-2001; 2001US-00016604.

PA (CHIR) CHIRON CORP.

PI Garcia P, Hardy SF, Williams LT, Escobedo J;

DR WPI; 2002-691475/74.

PT Novel isolated polypeptides useful for diagnosis of prostate cancer.

PS Disclosure; Page 129; 152pp; English.

CC The invention describes novel isolated polypeptides (I, Ib) useful for
 CC diagnosing prostate cancer comprising obtaining a patient sample
 CC containing prostate cells and detecting the presence or absence of an
 CC expression product of a HML-2 endogenous retrovirus in a patient sample.
 CC Polynucleotides associated with (I) are useful for diagnosis or treatment
 CC of testicular cancer, multiple sclerosis or insulin-dependent diabetes
 CC mellitus. An inhibitor of a HML-2 protease and a transdominant negative
 CC mutant of HML-2 CORF are also useful in the manufacture of a medicament
 CC for treating prostate cancer. (I) and (Ib) are useful for generating
 CC antibodies specific to the polypeptides associated with cancer, as
 CC targets for therapeutic intervention, and in immunising a transgenic
 CC animal. This sequence represents a region of human endogenous retrovirus
 CC (herv) of the HML-2 sub-group

SO Sequence 408 BP; 97 A; 113 C; 83 G; 115 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 6; Length 408;
 Best Local Similarity 100.0%; Pred. No. 5.9e-41;
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGGAAAAACCGCCTTAGGGCTGAGGTGGGACCTGCGGGACGAATACTGCTTT 60
DB 1 GAGATAGGGAAAAACCGCCTTAGGGCTGAGGTGGGACCTGCGGGACGAATACTGCTTT 60
QY 61 GTAAGCACTGAGATGTTTATGTGTATGCATATCTTAAAGCAGACACTTAATCCTTTAC 120
DB 61 GTAAGCACTGAGATGTTTATGTGTATGCATATCTTAAAGCAGACACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAGAACCCTTTGTTTAC 150
DB 121 ATTGCTATGATGCAAGAACCCTTTGTTTAC 150

RESULT 6
ADCl6257
ID ADCl6257 standard; DNA; 408 BP.
AC ADCl6257;
XX
XX
DT 18-DEC-2003 (first entry)
XX
DE Human herv-k(hml-2.hom) RUS region nucleotide sequence SEQ ID NO:4.
XX
XX prostate cancer; HML-2; env; cytosolic; neuroprotective; antidiabetic;
KM immunostimulant; vaccine; gene therapy; breast cancer; testicular cancer;
KM multiple sclerosis; insulin-dependent diabetes mellitus; human; gene; ds.
XX
XX Homo sapiens.
OS
XX WO2003050258-A2.
PN
XX 19-JUN-2003.
PD
XX 09-DEC-2002; 2002WO-US039344.
PF
XX 07-DEC-2001; 2001US-00016604.
PR 07-DEC-2001; 2001US-034064P.
PR 07-DEC-2001; 2001WO-US047824.
PR 12-JUN-2002; 2002US-0388046P.
XX
PA (CHIR) CHIRON CORP.
PI Garcia P, Hardy SF, Williams LT, Escobedo J;
XX WPI; 2003-577296/54.
DR
XX
PT Diagnosing prostate cancer, comprises detecting the presence or absence
PT of HML-2 (human endogenous retrovirus) expression product in a patient
PT sample.
XX
XX Disclousure; SEQ ID NO 4; 117pp; English.

PS The present invention describes a method for diagnosing prostate cancer,
XX comprising detecting the presence or absence of HML-2 expression product
XX in a patient sample, where the expression product is produced by a
XX cloning event in which the 5' region and start codon of the env coding
XX region are joined to a downstream coding region in the reading frame +2
XX relative to that of env. Also described: (1) an isolated polynucleotide;
XX (2) an isolated polypeptide; (3) an antibody that binds to the
XX polypeptide; and (4) an immunogenic composition. Prostate cancer related
XX sequences have cytosolic, neuroprotective, antidiabetic and
XX immunostimulant activities, and can be used in vaccines and in gene
XX therapy. A polynucleotide, polypeptide or antibody of the invention can
XX be used in the manufacture of a medicament for preventing or treating
XX prostate, breast or testicular cancer, multiple sclerosis or insulin-
XX dependent diabetes mellitus. The present sequence represents a sequence
XX which is used in the exemplification of the present invention.

XX
SQ Sequence 408 BP; 97 A; 113 C; 83 G; 115 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 9; Length 408;
Best Local Similarity 100.0%; Pred. No. 5.9e-41;

Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGATAGGGAAAAACCGCCTTAGGGCTGAGGTGGGACCTGCGGGACGAATACTGCTTT 60
DB 1 GAGATAGGGAAAAACCGCCTTAGGGCTGAGGTGGGACCTGCGGGACGAATACTGCTTT 60
QY 61 GTAAGCACTGAGATGTTTATGTGTATGCATATCTTAAAGCAGACACTTAATCCTTTAC 120
DB 61 GTAAGCACTGAGATGTTTATGTGTATGCATATCTTAAAGCAGACACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAGAACCCTTTGTTTAC 150
DB 121 ATTGCTATGATGCAAGAACCCTTTGTTTAC 150

RESULT 7
ABX04572
ID ABX04572 standard; DNA; 879 BP.
AC ABX04572;
XX
XX
DT 14-JAN-2003 (first entry)
XX
DE Human endogenous retrovirus k (herv-k) U3R region.
XX
XX Human; endogenous retrovirus; herv; prostate cancer; testicular cancer;
KM multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
KM cancer; transgenic animal; ds.
XX
XX Human endogenous retrovirus.
OS
XX WO200246477-A2.
PN
XX 13-JUN-2002.
PD
XX 07-DEC-2001; 2001WO-US047824.
PF
XX 07-DEC-2000; 2000US-0251830P.
PR 07-DEC-2001; 2001US-00016604.
PR
XX
PA (CHIR) CHIRON CORP.
PI Garcia P, Hardy SF, Williams LT, Escobedo J;
XX WPI; 2002-691475/74.
DR
XX
PT Novel isolated polypeptides useful for diagnosis of prostate cancer.
PT Claim 4; Page 129; 152pp; English.

PS The invention describes novel isolated polypeptides (I, Ib) useful for
XX diagnosing prostate cancer comprising obtaining a patient sample
XX containing prostate cells and detecting the presence or absence of an
XX expression product of a HML-2 endogenous retrovirus in a patient sample.
XX Polynucleotides associated with (I) are useful for diagnosis or treatment
XX of testicular cancer, multiple sclerosis or insulin-dependent diabetes
XX mellitus. An inhibitor of a HML-2 protease and a transdominant negative
XX mutant of HML-2 cORF are also useful in the manufacture of a medicament
XX for treating prostate cancer. (I) and (Ib) are useful for generating
XX antibodies specific to the polypeptides associated with cancer, as
XX targets for therapeutic intervention, and in immunising a transgenic
XX animal. This sequence represents a region of human endogenous retrovirus
XX (herv) of the HML-2 sub-group

XX
SQ Sequence 879 BP; 231 A; 212 C; 211 G; 225 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 6; Length 879;
Best Local Similarity 100.0%; Pred. No. 7.9e-41;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGATAGGGAAAAACCGCCTTAGGGCTGAGGTGGGACCTGCGGGACGAATACTGCTTT 60
DB 561 GAGATAGGGAAAAACCGCCTTAGGGCTGAGGTGGGACCTGCGGGACGAATACTGCTTT 620

Qy 61 GTAAAGCAGTGAAGTGTATGTATGATCATATCTAAAGACAGCACTTAATCCTTAC 120
Db 621 GTAAAGCAGTGAAGTGTATGTATGTATGATCATATCTAAAGACAGCACTTAATCCTTAC 680

Qy 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
Db 681 ATTGCTATGATGCAAGACCTTTGTTTAC 710

RESULT 8
ADCl6258
ID ADCl6258 standard; DNA; 879 BP.
XX
AC ADCl6258;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human herV-K(hml-2.hom) U3R region nucleotide sequence SEQ ID NO:5.
XX
KM prostate cancer; HML-2; env; cytosolic; neuroprotective; antidiabetic;
KW immunostimulant; vaccine; gene therapy; breast cancer; testicular cancer;
KW multiple sclerosis; insulin-dependent diabetes mellitus; human; gene; ds.
XX
OS Homo sapiens.
XX
PN MO2003050258-A2.
XX
PD 19-JUN-2003.
XX
PF 09-DEC-2002; 2002MO-US039344.
XX
PR 07-DEC-2001; 2001US-00016604.
PR 07-DEC-2001; 2001US-034064P.
PR 07-DEC-2001; 2001MO-US047824.
PR 12-JUN-2002; 2002US-0388046P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Garcia P, Hardy SF, Williams LT, Escobedo J;
XX
DR WPI; 2003-577296/54.
XX
PT Diagnosing prostate cancer, comprises detecting the presence or absence
PT of HML-2 (human endogenous retrovirus) expression product in a patient
PT sample.
XX
PS Disclosure; SEQ ID NO 5; 117bp; English.
XX
CC The present invention describes a method for diagnosing prostate cancer,
CC comprising detecting the presence or absence of HML-2 expression product
CC in a patient sample, where the expression product is produced by a
CC splicing event in which the 5' region and start codon of the env coding
CC region are joined to a downstream coding region in the reading frame +2
CC relative to that of env. Also described: (1) an isolated polynucleotide;
CC (2) an isolated polypeptide; (3) an antibody that binds to the
CC polypeptide; and (4) an immunogenic composition. Prostate cancer related
CC sequences have cytosolic, neuroprotective, antidiabetic and
CC immunostimulant activities, and can be used in vaccines and in gene
CC therapy. A polynucleotide, polypeptide or antibody of the invention can
CC be used in the manufacture of a medicament for preventing or treating
CC prostate, breast or testicular cancer, multiple sclerosis or insulin-
CC dependent diabetes mellitus. The present sequence represents a sequence
CC which is used in the exemplification of the present invention.
XX
SQ Sequence 879 BP; 231 A; 212 C; 211 G; 225 T; 0 U; 0 Other;

Query March 100.0%; Score 150; DB 9; Length 879;
Best Local Similarity 100.0%; Pred. No. 7.9e-41;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGATAGGAAAAAAGCCGCTTAGGGCTGAGAGTGGAGACTGCGGGAGCAATACCTT 60
Db 621 GTAAAGCAGTGAAGTGTATGTATGATCATATCTAAAGACAGCACTTAATCCTTAC 680

Db 561 GAGATAGGAAAAAAGCCGCTTAGGGCTGAGAGTGGAGACTGCGGGAGCAATACCTT 620
Qy 61 GTAAAGCAGTGAAGTGTATGTATGATCATATCTAAAGACAGCACTTAATCCTTAC 120
Db 621 GTAAAGCAGTGAAGTGTATGTATGATCATATCTAAAGACAGCACTTAATCCTTAC 680

Qy 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
Db 681 ATTGCTATGATGCAAGACCTTTGTTTAC 710

RESULT 9
ABX04691
ID ABX04691 standard; DNA; 968 BP.
XX
AC ABX04691;
XX
DT 14-JAN-2003 (first entry)
XX
DE Human endogenous retrovirus k (herV-K) long terminal repeat #1.
XX
KM Human endogenous retrovirus; herV; prostate cancer; testicular cancer;
KW multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
KW cancer; transgenic animal; ds.
XX
OS Human endogenous retrovirus.
XX
PN MO200246477-A2.
XX
PD 13-JUN-2002.
XX
PF 07-DEC-2001; 2001MO-US047824.
XX
PR 07-DEC-2000; 2000US-0251830P.
PR 07-DEC-2001; 2001US-00016604.
XX
PA (CHIR) CHIRON CORP.
XX
PI Garcia P, Hardy SF, Williams LT, Escobedo J;
XX
DR WPI; 2002-691475/74.
XX
PT Novel isolated polypeptides useful for diagnosis of prostate cancer.
XX
PS Claim 18; Page 148; 152pp; English.
XX
CC The invention describes novel isolated polypeptides (I, Ib) useful for
CC diagnosing prostate cancer comprising obtaining a patient sample
CC containing prostate cells and detecting the presence or absence of an
CC expression product of a HML-2 endogenous retrovirus in a patient sample.
CC Polynucleotides associated with (I) are useful for diagnosis or treatment
CC of testicular cancer, multiple sclerosis or insulin-dependent diabetes
CC mellitus. An inhibitor of a HML-2 protease and a transdominant negative
CC mutant of HML-2 CORF are also useful in the manufacture of a medicament
CC for treating prostate cancer. (I) and (Ib) are useful for generating
CC antibodies specific to the polypeptides associated with cancer, as
CC targets for therapeutic intervention, and in immunising a transgenic
CC animal. This sequence represents a region of human endogenous retrovirus
CC (herV) of the HML-2 sub-group
XX
SQ Sequence 968 BP; 249 A; 242 C; 225 G; 252 T; 0 U; 0 Other;

Query March 100.0%; Score 150; DB 6; Length 968;
Best Local Similarity 100.0%; Pred. No. 8.2e-41;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGATAGGAAAAAAGCCGCTTAGGGCTGAGAGTGGAGACTGCGGGAGCAATACCTT 60
Db 561 GAGATAGGAAAAAAGCCGCTTAGGGCTGAGAGTGGAGACTGCGGGAGCAATACCTT 620

Qy 61 GTAAAGCAGTGAAGTGTATGTATGATCATATCTAAAGACAGCACTTAATCCTTAC 120
Db 621 GTAAAGCAGTGAAGTGTATGTATGATCATATCTAAAGACAGCACTTAATCCTTAC 680

Db 1670 ATTGCTATGATGCAGAACCTTGTTCAC 1639

RESULT 12

AAK70301/C

ID AAK70301 standard; DNA; 6556 BP.

XX

AC AAK70301;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25113.

XX

KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX

OS cytostatic; gene therapy; vaccine; metastasis; ds.

XX

OS Homo sapiens.

XX

PN WO200157182-A2.

XX

PD 09-AUG-2001.

XX

PF 17-JAN-2001, 2001WO-US001354.

XX

PR 31-JAN-2000, 2000US-0179065P.

PR 04-FEB-2000, 2000US-0180628P.

PR 24-FEB-2000, 2000US-0184664P.

PR 02-MAR-2000, 2000US-0186350P.

PR 16-MAR-2000, 2000US-0189874P.

PR 17-MAR-2000, 2000US-0190076P.

PR 18-APR-2000, 2000US-0198123P.

PR 19-MAY-2000, 2000US-0205515P.

PR 07-JUN-2000, 2000US-0209467P.

PR 28-JUN-2000, 2000US-0214886P.

PR 30-JUN-2000, 2000US-0215135P.

PR 07-JUL-2000, 2000US-0216647P.

PR 11-JUL-2000, 2000US-0217487P.

PR 11-JUL-2000, 2000US-0217487P.

PR 14-JUL-2000, 2000US-0218290P.

PR 26-JUL-2000, 2000US-0220963P.

PR 26-JUL-2000, 2000US-0220964P.

PR 14-AUG-2000, 2000US-0224518P.

PR 14-AUG-2000, 2000US-0224519P.

PR 14-AUG-2000, 2000US-0225213P.

PR 14-AUG-2000, 2000US-0225214P.

PR 14-AUG-2000, 2000US-0225266P.

PR 14-AUG-2000, 2000US-0225267P.

PR 14-AUG-2000, 2000US-0225268P.

PR 14-AUG-2000, 2000US-0225270P.

PR 14-AUG-2000, 2000US-0225447P.

PR 14-AUG-2000, 2000US-0225757P.

PR 14-AUG-2000, 2000US-0225758P.

PR 14-AUG-2000, 2000US-0225759P.

PR 18-AUG-2000, 2000US-0226279P.

PR 22-AUG-2000, 2000US-0226681P.

PR 22-AUG-2000, 2000US-0226868P.

PR 22-AUG-2000, 2000US-0227182P.

PR 23-AUG-2000, 2000US-0227009P.

PR 30-AUG-2000, 2000US-0228924P.

PR 01-SEP-2000, 2000US-0229287P.

PR 01-SEP-2000, 2000US-0229343P.

PR 01-SEP-2000, 2000US-0229344P.

PR 01-SEP-2000, 2000US-0229345P.

PR 05-SEP-2000, 2000US-0229509P.

PR 05-SEP-2000, 2000US-0229513P.

PR 06-SEP-2000, 2000US-0230437P.

PR 06-SEP-2000, 2000US-0230438P.

PR 08-SEP-2000, 2000US-0231242P.

PR 08-SEP-2000, 2000US-0231243P.

PR 08-SEP-2000, 2000US-0231244P.

PR 08-SEP-2000, 2000US-0231413P.

PR 08-SEP-2000, 2000US-0231414P.

PR 08-SEP-2000, 2000US-0232080P.

PR 08-SEP-2000, 2000US-0232081P.

PR 12-SEP-2000, 2000US-0231968P.

PR 14-SEP-2000, 2000US-0232397P.

PR 14-SEP-2000, 2000US-0232398P.

PR 14-SEP-2000, 2000US-0232399P.

PR 14-SEP-2000, 2000US-0232400P.

PR 14-SEP-2000, 2000US-0232401P.

PR 14-SEP-2000, 2000US-0233063P.

PR 14-SEP-2000, 2000US-0233064P.

PR 14-SEP-2000, 2000US-0233065P.

PR 21-SEP-2000, 2000US-0234223P.

PR 21-SEP-2000, 2000US-0234274P.

PR 25-SEP-2000, 2000US-0234997P.

PR 25-SEP-2000, 2000US-0234998P.

PR 26-SEP-2000, 2000US-0235484P.

PR 27-SEP-2000, 2000US-0235834P.

PR 27-SEP-2000, 2000US-0235836P.

PR 29-SEP-2000, 2000US-0236372P.

PR 29-SEP-2000, 2000US-0236373P.

PR 29-SEP-2000, 2000US-0236367P.

PR 29-SEP-2000, 2000US-0236368P.

PR 29-SEP-2000, 2000US-0236369P.

PR 29-SEP-2000, 2000US-0236370P.

PR 02-OCT-2000, 2000US-0237037P.

PR 02-OCT-2000, 2000US-0237038P.

PR 02-OCT-2000, 2000US-0237039P.

PR 02-OCT-2000, 2000US-0237040P.

PR 13-OCT-2000, 2000US-0239935P.

PR 13-OCT-2000, 2000US-0239937P.

PR 20-OCT-2000, 2000US-0240960P.

PR 20-OCT-2000, 2000US-0241221P.

PR 20-OCT-2000, 2000US-0241785P.

PR 20-OCT-2000, 2000US-0241786P.

PR 20-OCT-2000, 2000US-0241787P.

PR 20-OCT-2000, 2000US-0241808P.

PR 20-OCT-2000, 2000US-0241809P.

PR 20-OCT-2000, 2000US-0241826P.

PR 01-NOV-2000, 2000US-0244617P.

PR 08-NOV-2000, 2000US-0246474P.

PR 08-NOV-2000, 2000US-0246475P.

PR 08-NOV-2000, 2000US-0246476P.

PR 08-NOV-2000, 2000US-0246477P.

PR 08-NOV-2000, 2000US-0246478P.

PR 08-NOV-2000, 2000US-0246523P.

PR 08-NOV-2000, 2000US-0246524P.

PR 08-NOV-2000, 2000US-0246525P.

PR 08-NOV-2000, 2000US-0246526P.

PR 08-NOV-2000, 2000US-0246527P.

PR 08-NOV-2000, 2000US-0246528P.

PR 08-NOV-2000, 2000US-0246529P.

PR 08-NOV-2000, 2000US-0246609P.

PR 08-NOV-2000, 2000US-0246610P.

PR 08-NOV-2000, 2000US-0246611P.

PR 08-NOV-2000, 2000US-0246613P.

PR 17-NOV-2000, 2000US-0249207P.

PR 17-NOV-2000, 2000US-0249208P.

PR 17-NOV-2000, 2000US-0249209P.

PR 17-NOV-2000, 2000US-0249210P.

PR 17-NOV-2000, 2000US-0249211P.

PR 17-NOV-2000, 2000US-0249212P.

PR 17-NOV-2000, 2000US-0249213P.

PR 17-NOV-2000, 2000US-0249214P.

PR 17-NOV-2000, 2000US-0249215P.

PR 17-NOV-2000, 2000US-0249216P.

PR 17-NOV-2000, 2000US-0249217P.

PR 17-NOV-2000, 2000US-0249218P.

PR 17-NOV-2000, 2000US-0249244P.

PR 17-NOV-2000, 2000US-0249245P.

PR 17-NOV-2000, 2000US-0249246P.

PR 17-NOV-2000, 2000US-0249256P.

PR 17-NOV-2000, 2000US-0249297P.

PR 17-NOV-2000, 2000US-0249299P.

XX 14-AUG-2002 (first entry)
DT Human cDNA differentially expressed in granulocytic cells #920.
XX
XX
XX Human; 89; granulocytic cell; DNA chip; bacterial infection;
KM viral infection; parasitic infection; protozoal infection;
KM fungal infection; sterile inflammatory disease; psoriasis;
KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KM cardiac reperfusion injury; renal reperfusion injury; AIDS;
KM adult respiratory distress syndrome; inflammatory bowel disease;
KM Crohn's disease; ulcerative colitis; periodontal disease;
KM granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
PN WO200228999-A2.
PD 11-APR-2002.
PF 03-OCT-2001; 2001WO-US030821.
PR 03-OCT-2000; 2000US-0237189P.
PA (GENE-) GENE LOGIC INC.
PI Beazer-Barclay Y, Weiseman SM, Yamaga S, Vockley J;
XX WPI; 2002-435328/46.
XX
XX Detecting granulocyte activation by detecting differential expression of
PT gene associated with granulocyte activation, which serves as diagnostic
PT markers that is useful for monitoring disease states and drug toxicity.
PS Claim 1; SEQ ID NO 920; 114pp; English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing the
CC expression level to an expression level in an unactivated GC, where
CC differential expression of Gs is indicative of GCA. Also included are
CC modulating (M2) GA by contacting GC with an agent that alters the
CC expression of at least one gene in Gs; (2) screening (M3) for an agent
CC capable of modulating GCA or an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease using the gene expression
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease, by detecting the level of
CC expression in a sample of the tissue of gene(s) from Gs, where the level
CC of expression of the gene is indicative of inflammation; (4) treating
CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease, by contacting a tissue having inflammation with an
CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC for screening an agent capable of modulating GCA preferably in an
CC inflammation in a tissue; M4 is useful for detecting an inflammation
CC (especially chronic) in a tissue, an allergic response in a subject,
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC cardiac reperfusion injury, renal reperfusion injury, AIDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease; also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
CC conditions. The present sequence represents a gene differentially
CC expressed in granulocytes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 222930 BP; 53286 A; 56778 C; 57221 G; 55643 T; 0 U; 2 Other;

Query Match 98.9%; Score 148.4; DB 6; Length 222930;
Best Local Similarity 99.3%; Pred. No. 2,2e-39;
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GAGATAGGAAAAAACCCTTAGGCTGAGAGTGGACCTGCGGACGAACTACTGCTTT 60
Db 98271 GAGATAGGAAAAAACCCTTAGGCTGAGAGTGGACCTGCGGACGAACTACTGCTTT 98212
Qy 61 GTAAAGCAGTCGAGATGTTTATGTATGATCATATTTAAAGACAGCACTTAATCTTTAC 120
Db 98211 GTAAAGCAGTCGAGATGTTTATGTATGATCATATTTAAAGACAGCACTTAATCTTTAC 98152
Qy 121 ATTGCTATGATGCAAGACCTTGTTCAC 150
Db 98151 ATTGCTATGATGCAAGACCTTGTTCAC 98122
RESULT 15
ABX04695
ID ABX04695 standard; DNA; 968 BP.
XX
XX ABX04695;
AC
XX
DT 14-JAN-2003 (first entry)
XX
XX Human endogenous retrovirus k (herv-k) long terminal repeat #5.
XX
XX Human; endogenous retrovirus; herv; prostate cancer; testicular cancer;
KM multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
KM cancer; transgenic animal; ds-
XX
XX Human endogenous retrovirus.
OS
PN WO200246477-A2.
PD 13-JUN-2002.
PF 07-DEC-2001; 2001WO-US047824.
PR 07-DEC-2000; 2000US-0251830P.
PR 07-DEC-2001; 2001US-00016604.
PA (CHIR) CHIRON CORP.
XX
XX Garcia P, Hardy SF, Williams LT, Escobedo J;
PI WPI; 2002-691475/74.
DR
XX
XX Novel isolated polypeptides useful for diagnosis of prostate cancer.
PT
XX
PS Claim 18; Page 149; 152pp; English.
XX
XX The invention describes novel isolated polypeptides (I, Ib) useful for
CC diagnosing prostate cancer comprising obtaining a patient sample
CC containing prostate cells and detecting the presence or absence of an
CC expression product of a HML-2 endogenous retrovirus in a patient sample.
CC Polynucleotides associated with (I) are useful for diagnosis or treatment
CC of testicular cancer, multiple sclerosis or insulin-dependent diabetes
CC mellitus. An inhibitor of a HML-2 protease and a transdominant negative
CC mutant of HML-2 CORP are also useful in the manufacture of a medicament
CC for treating prostate cancer. (I) and (Ib) are useful for generating
CC antibodies specific to the polypeptides associated with cancer, as
CC targets for therapeutic intervention, and in immunising a transgenic
CC animal. This sequence represents a region of human endogenous retrovirus
CC (herv) of the HML-2 sub-group
XX
XX Sequence 968 BP; 250 A; 237 C; 225 G; 256 T; 0 U; 0 Other;

Query Match 97.9%; Score 146.8; DB 6; Length 968;
Best Local Similarity 98.7%; Pred. No. 1e-39;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY      1 GAGATAGGGAAAAACCGCTTTAGGGCTGAGGTGGACCTGCGGGCAGCAATACTGCTTT 60
      |||
Db      561 GAGATAGGGAAAAACCGCTTTAGGGCTGAGGTGGACCTGCGGGCAGCAATACTGCTTT 620
      |||
QY      61  GTAAGCACTGAGATGTTTATGTGTATGCATATCTAAAGCAGCACTTAATCCTTAC 120
      |||
Db      621 GTAAGCATTGAGATGTTTATGTGTATGCATATCTAAAGCAGCACTTAATCCTTAC 680
      |||
QY      121 ATTGCTATGATGCAAAAGACCTTTGTCAC 150
      |||
Db      681 CTTGCTATGATGCAAAAGACCTTTGTCAC 710
      |||
```

Search completed: February 27, 2004, 22:44:04
Job time : 79.9883 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 27, 2004, 22:29:18 ; Search time 609.038 Seconds
(without alignments)
7354.759 Million cell updates/sec

Title: US-10-016-604-155

Perfect score: 150
Sequence: 1 gagatagggaacacgcctc.....atgcaagaccttcttcac 150

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: em_estbda:*
2: em_estbda:*
3: em_estbda:*
4: em_estbda:*
5: em_estbda:*
6: em_estbda:*
7: em_estbda:*
8: em_estbda:*
9: em_estbda:*
10: em_estbda:*
11: em_estbda:*
12: em_estbda:*
13: em_estbda:*
14: em_estbda:*
15: em_estbda:*
16: em_estbda:*
17: em_estbda:*
18: em_estbda:*
19: em_estbda:*
20: em_estbda:*
21: em_estbda:*
22: em_estbda:*
23: em_estbda:*
24: em_estbda:*
25: em_estbda:*
26: em_estbda:*
27: em_estbda:*
28: em_estbda:*
29: em_estbda:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	148.4	98.9	444	9	AI066622_0282b02.x
C 2	148.4	98.9	1204	12	BM907033_AGENCOURT
C 3	147.4	98.3	460	14	H80802_Y964d11.r1
C 4	146.8	97.9	356	10	BF737628_PMI-KT003

Result No.	Score	Query Match	Length	ID	Description
C 5	146.8	97.9	448	9	AA393030_z69f06.x
C 6	146.8	97.9	488	12	BI048071_PMI-ST030
C 7	146.8	97.9	524	12	BM693416_UI-E-CK1
C 8	146.8	97.9	576	13	BU632008_UI-H-FE1
C 9	146.8	97.9	633	13	BM668693_UI-E-CK1
C 10	146.8	97.9	689	13	BQ575726_UI-H-EZ1
C 11	146.8	97.9	706	13	BU618928_UI-H-FH1
C 12	146.8	97.9	707	12	BM981862_UI-CF-EN1
C 13	145.2	96.8	438	10	BF918605_CMO-NT013
C 14	145.2	96.8	704	12	BM989989_UI-H-D10
C 15	145.2	96.8	818	13	BQ216036_AGENCOURT
C 16	143.6	95.7	452	9	AA573159_nms1e12.s
C 17	143.6	95.7	1698	11	BC036118_Homo sapi
C 18	142	94.7	874	14	CD558661_AGENCOURT
C 19	140.4	93.6	372	13	BQ360083_CMO-HN020
C 20	140.4	93.6	586	9	AA708837_zf97d06.s
C 21	140.4	93.6	689	14	CD684466_EST1006.h
C 22	140.4	93.6	798	12	BI758163_603023895
C 23	138.8	92.5	409	13	BQ312183_QV3-BN037
C 24	137.2	91.5	649	9	AL049055_DKFZP434P
C 25	136.4	90.9	541	14	W52520_RCS-NT026
C 26	135.6	90.4	351	12	BI038269_AA703380_zj12b06.s
C 27	135.6	90.4	471	9	AA703380_UI-H-ED0
C 28	135.6	90.4	570	12	BQ027101_UI-H-E11
C 29	135.6	90.4	670	12	BQ006641_UI-H-E11
C 30	135.6	90.4	897	13	BX374012_BX374012
C 31	135.6	90.4	954	13	BX327468_BX327468
C 32	135.2	90.1	232	10	BF948093_MK3-NN022
C 33	135.2	90.1	322	14	T05809_EST03698.Fe
C 34	133.8	89.2	292	9	AA076809_7B03C10.C
C 35	133.6	89.1	975	9	AL565978_AL565978
C 36	133.2	88.8	485	12	BM667549_UI-E-DX0
C 37	133.2	88.8	678	14	CA449778_UI-H-E11
C 38	131.8	87.9	305	13	BQ318376_CMA-CT051
C 39	131.8	87.9	329	13	BQ318375_CMA-CT051
C 40	131.8	87.9	670	14	CB850840_UI-CF-EN1
C 41	131.4	87.6	300	9	AA716377_zg62b12.s
C 42	130.4	86.9	339	9	AA633703_Ag87b07.s
C 43	130.2	86.8	622	12	BM985460_UI-CF-EC1
C 44	130	86.7	632	10	AW963304_EST375377
C 45	127.6	85.1	376	13	BQ318373_CMA-CT051

ALIGNMENTS

RESULT 1
LOCUS AI066622/c 444 bp mRNA linear EST 01-OCT-1998
DEFINITION 0282b02.x1 Soares senescent fibroblasts NBHSF Homo sapiens CDNA
clone IMAGE:1681803 3' similar to contains LTR5.tl LTR5 repetitive
element ;, mRNA sequence.
ACCESSION AI066622 GI:3367324
VERSION AI066622
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 444)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapdb-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 615 Std Error: 0.00
Seq primer: -40m13 fwd: RT from Amersham.
Location/Qualifiers
1..444
/organism="Homo sapiens"


```

/db xref="taxon:9606"
/dev strage="Adult"
/clone_lib="RT0037"
/note="Origin: Bladder tumor; Vector: puc19; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORSTS PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 19 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

Query Match	97.9%;	Score 146.8;	DB 10;	Length 356;
Best Local Similarity	98.7%;	Pred. No. 4.3e-35;		
Matches 148; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046
047
048
049
050
051
052
053
054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100

61 GTAAGCACTGAGTGTATGTATGCATATCTAAAGCACAGACCTTATCCTTTAC 120

QY 121 ATTGCTATGATGCAAGACCTTTGTTAC 150
|||||
|||

RESULT 5

DEFINITION 2696f06.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:272619 5' similar to contains LTR5.b2 LTR5 repetitive element ;, mRNA

ACCESSION AA393030
VERSION AA393030.1
KEYWORDS
EST. GI: 2046072

REFERENCE

Eukaryota: Metazoa: Chordata: Craniata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo. 1 (bases 1 to 448)

Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellendberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.

JOURNAL COMMENT
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

IMAGE Consortium (info@image.jnl.gov) for further information.
Insert Length: 663 Std Error: 0.00
Seq primer: -28ml3 rev2 ET from Amersham

```

FEATURES
  source      location/Qualifiers
1.  .448     /organism="Homo sapiens"

```

```
/db_xref="GDB:5924530"
/db_xref="Eaxom:9606"
/clone="IMAGE:727619"
```

```

/lab_host="DH10B"
/clone_lib="Soares testis_NHT"
/notes="Vector: pRTT3D-Pac (Pharmacia) with a modified

```

was prepared from mRNA obtained from Clontech

Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTACCAATCGAAGTGGAGGCGCCCAATTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to C₀5, and was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 97.9%; Score 146.8; DB 9; Length 448;
Best Local Similarity 98.7%; Pred. No. 4.6e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAACCCCTTAGGCTGAGTGGACCTGCGGACGACAACTACTGCTTT 60
DB 209 GAGATAGGAAAAACCCCTTAGGCTGAGTGGACCTGCGGACGACAACTACTGCTTT 268
QY 61 GTAAGACACTGAGATGTTTATGTATGCAATCTAAAGACAGACCTTAATCCTTTAC 120
DB 269 GTAAGACACTGAGATGTTTATGTATGCAATCTAAAGACAGACCTTAATCCTTTAC 328
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 329 ATTGCTATGATGCAAGACCTTTGTTTAC 358

RESULT 6
BI048071/c 488 bp mRNA linear EST 14-JUN-2001
LOCUS BI048071
DEFINITION PM2-ST0303-230301-011-h03 ST0303 Homo sapiens cDNA, mRNA sequence.
ACCESSION BI048071
VERSION BI048071.1 GI:14454693
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongsenel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE
JOURNAL MEDLINE
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LIC Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM2&ct=PM2-ST0303-230301-011-h03&ct3=2001-03-23&ct4=1)
Seq primer: puc 18 forward
High quality sequence start: 7
High quality sequence stop: 488.
Location/Qualifiers

FEATURES

source
1. .488
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="adult"
/clone_id="ST0303"
/note="Organ: stomach; Vector: puc18; site_1: Sma1;

Site_2: Sma1; A mini-library was made by cloning products derived from ONESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 97.9%; Score 146.8; DB 12; Length 488;
Best Local Similarity 98.7%; Pred. No. 4.8e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAACCCCTTAGGCTGAGTGGACCTGCGGACGACAACTACTGCTTT 60
DB 232 GAGATAGGAAAAACCCCTTAGGCTGAGTGGACCTGCGGACGACAACTACTGCTTT 173
QY 61 GTAAGACACTGAGATGTTTATGTATGCAATCTAAAGACAGACCTTAATCCTTTAC 120
DB 172 GTAAGACACTGAGATGTTTATGTATGCAATCTAAAGACAGACCTTAATCCTTTAC 113
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 112 ATTGCTATGATGCAAGACCTTTGTTTAC 83

RESULT 7
BM693416 524 bp mRNA linear EST 28-FEB-2002
LOCUS BM693416
DEFINITION UI-E-CK1-afm-b-10-0-UI-r1 UI-E-CK1 Homo sapiens cDNA clone
ACCESSION UI-E-CK1-afm-b-10-0-UI 5', mRNA sequence.
VERSION BM693416.1 GI:19006674
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL MEDLINE
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.regen.com).
The following repetitive elements were found in this cDNA
sequence: 1-521, >LTR5#LTR/Retroviral
Seq primer: M13 Reverse.
Location/Qualifiers

FEATURES

source
1. .524
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-CK1-afm-b-10-0-UI"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="PH10B (Life Technologies) (71 phage resistant)"
/clone_lib="UI-E-CK1"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-CK1 is a normalized cDNA library containing the

following tissue(s): Retina Foveal and Macular. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Query Match 97.9%; Score 146.8; DB 12; Length 524;
Best Local Similarity 98.7%; Pred. No. 4.9e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAACCGCTTAGGGCTGAGGTGGAGACCTGCGGAGCAATACCTGCTT 60
DB 301 GAGATAGGAAAAACCGCTTAGGGCTGAGGTGGAGACCTGCGGAGCAATACCTGCTT 360
QY 61 GTAAGCACTGAGATGTTTATGTGTATGATATCTAAAGCAGACCTTAATCCTTTAC 120
DB 361 GTAAGCACTGAGATGTTTATGTGTATGATATCTAAAGCAGACCTTAATCCTTTAC 420
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 421 ATTGCTATGATGCAAGACCTTTGTTTAC 450

RESULT 8

BU632008/c 576 bp mRNA linear EST 23-SEP-2002
LOCUS UI-H-FEI-bea-c-09-0-UI.81 NCI CGAP FEI Homo sapiens CDNA clone
DEFINITION UI-H-FEI-bea-c-09-0-UI 3', mRNA sequence.
ACCESSION BU632008
VERSION BU632008.1 GI:23299263
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 576)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: James Martin
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this CDNA
sequence: 18-576, >LTR5#LTR/Retroviral (matched complement)
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES

source location/Qualifiers
1. 576
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-FEI-bea-c-09-0-UI"
/tissue_type="Cell lines"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP FEI"
/note="Organ: Chondrosarcoma; Vector: p773-Pac
(Pharmacia) with a modified polylinker; Site_1: EcoR I;

Site 2: Not I; NCI CGAP FEI is a normalized cDNA library derived from a pool of mRNA obtained from 3 cell lines from grade II chondrosarcoma tissues. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGCTACGAGC. The cell lines were provided by Dr James Martin from the University of Iowa.
TAG_TISSUE=Human grade 2 chondrosarcoma cell line pool
TAG_LIB=UI-H-FEI
TAG_SEQ=CGCTACGAGC"

ORIGIN

Query Match 97.9%; Score 146.8; DB 13; Length 576;
Best Local Similarity 98.7%; Pred. No. 5.1e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAACCGCTTAGGGCTGAGGTGGAGACCTGCGGAGCAATACCTGCTT 60
DB 334 GAGATAGGAAAAACCGCTTAGGGCTGAGGTGGAGACCTGCGGAGCAATACCTGCTT 275
QY 61 GTAAGCACTGAGATGTTTATGTGTATGATATCTAAAGCAGACCTTAATCCTTTAC 120
DB 274 GTAAGCACTGAGATGTTTATGTGTATGATATCTAAAGCAGACCTTAATCCTTTAC 215
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 214 ATTGCTATGATGCAAGACCTTTGTTTAC 185

RESULT 9

BM668693/c 633 bp mRNA linear EST 27-FEB-2002
LOCUS UI-E-CK1-afm-b-10-0-UI.82 UI-E-CK1 Homo sapiens CDNA clone
DEFINITION UI-E-CK1-afm-b-10-0-UI 3', mRNA sequence.
ACCESSION BM668693
VERSION BM668693.1 GI:18978590
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 633)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
79044477
8889548

JOURNAL COMMENT

Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this CDNA
sequence: 18-633, >LTR5#LTR/Retroviral (matched complement)
Seq primer: M13 Forward
POLYA=yes.
location/Qualifiers

FEATURES

source

1. .633
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-CK1-afm-b-10-0-UI"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_1fb="UI-E-CK1"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: Ecor I; Site 2: Not I; UI-E-CK1 is a normalized cDNA library containing the following tissue(s): Retina Foveal and Macular. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).
TAG_TISSUE=Foveal and Macular Retina
TAG_LIB=UI-E-CK1
TAG_SEQ=GTC"

ORIGIN

Query Match 97.9%; Score 146.8; DB 12; Length 633;
Best Local Similarity 98.7%; Pred. No. 5.2e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGATGGGAAAAACCGCCTTAGGGCTGGAGTGGGACCTGGGCGAGCAATCTGCTT 60
DB 334 GAGATGAGGAAAAACCGCCTTAGGGCTGGAGTGGGACCTGGGCGAGCAATCTGCTT 275
QY 61 GTAAGCACTGAGATGTTATGTTATGATGATCTAATCAAGACAGCACTTAATCCTTAC 120
DB 274 GTAAGCACTGAGATGTTATGTTATGATGATCTAATCAAGACAGCACTTAATCCTTAC 215
QY 121 ATTGCTATGATGCAAGACCTTTGTTAC 150
DB 214 ATTGCTATGATGCAAGACCTTTGTTAC 185

RESULT 10
B0575726/c 689 bp mRNA linear EST 19-JUN-2002
LOCUS
DEFINITION
UI-H-EZ1-bbg-f-06-0-UI.s1 NCI CGAP Ch2 Homo sapiens cDNA clone
B0575726
B0575726.1 GI:21479043
VERSION
B0575726.1
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 689)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Dr. Steven Gitelis/ Rush Presbyterian, Dept. of
Orthopaedics
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu

The following repetitive elements were found in this cDNA
sequence: 16-689, >LTR5#LTR/Retroviral (matched complement)
Seq primer: M13 FORWARD
POLVA=Yes.

FEATURES
source

Location/Qualifiers

1. .689
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-EZ1-bbg-f-06-0-UI"
/tissue_type="Chondrosarcoma Grade II"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_1fb="NCI CGAP Ch2"
/note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: Ecor I; Site 2: Not I; NCI CGAP Ch2 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma Grade II. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is TGATCAGCT.
TAG_TISSUE=grade-2-chondrosarcoma
TAG_LIB=UI-H-EZ1
TAG_SEQ=ATCTAATATG"

ORIGIN

Query Match 97.9%; Score 146.8; DB 13; Length 689;
Best Local Similarity 98.7%; Pred. No. 5.4e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGATGGGAAAAACCGCCTTAGGGCTGGAGTGGGACCTGGGCGAGCAATCTGCTT 60
DB 271 GAGATGAGGAAAAACCGCCTTAGGGCTGGAGTGGGACCTGGGCGAGCAATCTGCTT 212
QY 61 GTAAGCACTGAGATGTTATGTTATGATGATCTAATCAAGACAGCACTTAATCCTTAC 120
DB 211 GTAAGCACTGAGATGTTATGTTATGATGATCTAATCAAGACAGCACTTAATCCTTAC 152
QY 121 ATTGCTATGATGCAAGACCTTTGTTAC 150
DB 151 ATTGCTATGATGCAAGACCTTTGTTAC 122

RESULT 11
B0618928/c 706 bp mRNA linear EST 23-SEP-2002
LOCUS
DEFINITION
UI-H-FH1-bf1-b-11-0-UI.s1 NCI CGAP FH1 Homo sapiens cDNA clone
B0618928
B0618928.1 GI:23285143
VERSION
B0618928
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 706)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: James Martin
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu. The following repetitive elements were found in this cDNA sequence: 18-706, >LTR5#LTR/retroviral (matched complement). Seq primer: M13 FORWARD
POLYA=yes.

FEATURES
source

Location/Qualifiers
1..706
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-FH1-bf1-b-11-0-UI"
/issue_type="Cell Line"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP_FH1"
/note="Organ: Chondrosarcoma; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP FH1 is a normalized cDNA library obtained from a cell line derived from grade I chondrosarcoma tissue. The library was constructed and normalized according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AGATCCGGC. The cell line was provided by Dr. James Martin from the University of Iowa.
TAG_TISSUE=Human Chondrosarcoma Cell Line C98 - Grade 1 Chondrosarcoma
TAG_LIB=UI-H-FH1
TAG_SEQ=AGATCCGGC"

ORIGIN

Query Match 97.9%; Score 146.8; DB 13; Length 706;
Best Local Similarity 98.7%; Pred. No. 5.5e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGTATAGGAAAAACCGCTTAGGGCTGAGGTGGAGACTCGGGACGACATATGCTT 60
DB 272 GAGTATAGGAAAAACCGCTTAGGGCTGAGGTGGAGACTCGGGACGACATATGCTT 213
QY 61 GTAAGCACTGAGATGTTTATGTATGATCATATCTAAAGCAGACACTTAATCCTTAC 120
DB 212 GTAAGCACTGAGATGTTTATGTATGATCATATCTAAAGCAGACACTTAATCCTTAC 153
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 152 ATTGCTATGATGCAAGACCTTTGTTTAC 123

RESULT 12
BM981862/c 707 bp mRNA linear EST 21-FEB-2003
LOCUS UI-CF-EN1-adv-g-12-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
DEFINITION BM981862
ACCESSION BM981862.1 GI:19604781
VERSION
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 707)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477

PUBMED 8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.research.com) or from Open Biosystems (www.openbiosystems.com).
The following repetitive elements were found in this cDNA sequence: 17-707, >LTR5#LTR/retroviral (matched complement). Seq primer: M13 FORWARD
POLYA=yes.

FEATURES
source

Location/Qualifiers
1..707
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EN1-adv-g-12-0-UI"
/issue_type="Primary Lung Cystic Fibrosis Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EN1"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-EN1 is a normalized cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT.
TAG_TISSUE=Human Lung Epithelial Cell lines untreated LPS 6hr to LPS 24h
TAG_LIB=UI-CF-EN1
TAG_SEQ=CTGCTCAGGT"

ORIGIN

Query Match 97.9%; Score 146.8; DB 12; Length 707;
Best Local Similarity 98.7%; Pred. No. 5.5e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGTATAGGAAAAACCGCTTAGGGCTGAGGTGGAGACTCGGGACGACATATGCTT 60
DB 298 GAGTATAGGAAAAACCGCTTAGGGCTGAGGTGGAGACTCGGGACGACATATGCTT 239
QY 61 GTAAGCACTGAGATGTTTATGTATGATCATATCTAAAGCAGACACTTAATCCTTAC 120
DB 238 GTAAGCACTGAGATGTTTATGTATGATCATATCTAAAGCAGACACTTAATCCTTAC 179
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 178 ATTGCTATGATGCAAGACCTTTGTTTAC 149

RESULT 13
BF918605 438 bp mRNA linear EST 19-JAN-2001
LOCUS CMO-NT0133-131000-614-b04 NT0133 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF918605
ACCESSION BF918605.1 GI:12314493
VERSION

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 438)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W., Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PubMed 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPSP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CMO&t2=CMO-NT0133-131000-614-B04&t3=2000-10-13&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 3
High quality sequence stop: 438.
Location/Qualifiers
1. 438
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NT0133"
/note="Organ: nervous tumor; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORSTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
ORIGIN
Query Match 96.8%; Score 145.2; DB 10; Length 438;
Best Local Similarity 98.0%; Pred. No. 1.4e-34;
Matches 147; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GAGATTAGGAAAAACCGCTTAGGGCTGAGGTGGGACCTGGGGCAATATGCTTT 60
DB 267 GAGATTAGGAAAAACCGCTTAGGGCTGAGGTGGGACCTGGGGCAATATGCTTT 326
QY 61 GTAAGACATGAGATGTTATGTATGATGATCTAAAGACAGCACTTAATCCTTTAC 120
DB 327 GTAAGACATGAGATGTTATGTATGATGATCTAAAGACAGCACTTAATCCTTTAC 386
QY 121 ATTGCTATGATGCAAGACCTTTGTTAC 150
DB 387 ATTGCTATGATGCAAGACCTTTGTTAC 416

RESULT 14
BM989989 704 bp mRNA linear EST 17-JUN-2002
LOCUS BM989989
DEFINITION UI-H-DIO-ato-h-05-0-UI.s1 NCI CGAP_D10 Homo sapiens cDNA clone
IMAGE:5862196 3', mRNA sequence.
ACCESSION BM989989
VERSION BM989989.1 GI:19709378
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 704)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps@mail.nih.gov
Tissue Procurement: Dr. Jose Mercende
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: http://image.llnl.gov
The following repetitive elements were found in this cDNA sequence: 16-704, >LIR5#LTR/Retroviral (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
1. 704
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5862196"
/cisse_type="Lung Focal Fibrosis"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP_D10"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP D10 is a cDNA library containing the following tissue(s): A pool of lung focal fibrosis. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA, synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ATACGCGGTC.
TAG TRISUB=lung with fibrosis
TAG_LIB=UI-H-DIO
TAG_SEQ=ATACGCGGTC"
ORIGIN
Query Match 96.8%; Score 145.2; DB 12; Length 704;
Best Local Similarity 98.0%; Pred. No. 1.7e-34;
Matches 147; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GAGATTAGGAAAAACCGCTTAGGGCTGAGGTGGGACCTGGGGCAATATGCTTT 60
DB 270 GAGATTAGGAAAAACCGCTTAGGGCTGAGGTGGGACCTGGGGCAATATGCTTT 211
QY 61 GTAAGACATGAGATGTTATGTATGATGATCTAAAGACAGCACTTAATCCTTTAC 120
DB 210 GTAAGACATGAGATGTTATGTATGATGATCTAAAGACAGCACTTAATCCTTTAC 151
QY 121 ATTGCTATGATGCAAGACCTTTGTTAC 150
DB 150 CTGCTATGATGCAAGACCTTTGTTAC 121

RESULT 15
BQ216036 818 bp mRNA linear EST 02-MAY-2002
LOCUS BQ216036
DEFINITION AGENCOURT_7553694 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6046113
5', mRNA sequence.
ACCESSION BQ216036
KEYWORDS

VERSION BQ216036.1 GI:20397447
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Straubeberg, Ph.D.
 Email: cgaps-romail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM13291 row: 9 column: 10
 High quality sequence stop: 577.

FEATURES

Source

Location/Qualifiers

1..818

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6046113"

/issue_type="embryonal carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_id="NIH_MGC_92"

/note="Organ: testis; Vector: PCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 2.5 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

ORIGIN

Query Match

96.8%; Score 145.2; DB 13; Length 818;

Best Local Similarity 98.0%; Pred. No. 1.8e-34; Mismatches 3; Indels 0; Gaps 0;

Matches 147; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAACCGCTTAGGGCTGAGGTGGACCTGCGGACGACATATCTGCTT 60
 |||
 Db 156 GAGATAGGAAAAACCGCTTAGGGCTGAGGTGGACCTGCGGACGACATATCTGCTT 97
 |||
 QY 61 GTAAAGACCTGAGATGTTTATGTATGTCATATCTAAAGACACGACTTATCTTTAC 120
 |||
 Db 96 GTAAAGCATGAAATGTTTATGTATGTCATATCTAAAGACACGACTTATCTTTAC 37
 |||
 QY 121 ATTGCTATGATGCAAGACCTTGTTCAC 150
 |||
 Db 36 CTGTCTATGATGCAAGACCTTGTTCAC 7
 |||

Search completed: February 28, 2004, 01:07:17
 Job time : 613.038 secs

This Page Blank (USPO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2004, 22:34:13 ; Search time 18.2216 Seconds
(without alignments)
4568.356 Million cell updates/sec

Title: US-10-016-604-155

Perfect score: 150

Sequence: 1 gagatagggagaaacgcgcct.....atgcgaagaccttcttccac 150

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	146	97.3	2525	US-09-482-273-20	Sequence 20, Appl
2	115.6	77.1	246240	US-08-724-394A-20	Sequence 20, Appl
3	115.6	77.1	246240	US-08-724-394A-21	Sequence 21, Appl
4	115.6	77.1	246240	US-08-724-394A-22	Sequence 22, Appl
5	106.8	71.2	176373	US-09-128-155-17	Sequence 17, Appl
6	92.8	61.9	589	US-09-328-111-600	Sequence 600, App
7	76	50.7	9388	US-08-991-789A-141	Sequence 141, App
8	76	50.7	9388	US-09-062-451-141	Sequence 141, App
9	76	50.7	9388	US-09-598-326-141	Sequence 141, App
10	76	50.7	9388	US-09-289-198-141	Sequence 141, App
11	76	50.7	9388	US-09-428-755-141	Sequence 141, App
12	51.6	34.4	2622	US-09-904-615-45	Sequence 45, Appl
13	39.2	26.1	519	US-08-991-789A-240	Sequence 240, App
14	39.2	26.1	519	US-09-062-451-240	Sequence 240, App
15	39.2	26.1	519	US-09-598-326-240	Sequence 240, App
16	39.2	26.1	519	US-09-289-198-240	Sequence 240, App
17	39.2	26.1	519	US-09-428-755-240	Sequence 240, App
18	38.4	25.6	90541	US-09-759-359A-3	Sequence 3, Appl
19	33.8	22.5	1086	US-08-991-789A-7	Sequence 7, Appl
20	33.8	22.5	1086	US-09-062-451-7	Sequence 7, Appl
21	33.8	22.5	1086	US-09-598-326-7	Sequence 7, Appl
22	33.8	22.5	1086	US-09-289-198-7	Sequence 7, Appl
23	33.8	22.5	1086	US-09-428-755-7	Sequence 7, Appl
24	31.2	20.8	459	US-09-621-976-2273	Sequence 2273, Ap
25	29.8	19.9	6157	US-08-956-171E-184	Sequence 184, Appl
26	28.4	18.9	7163	US-08-961-527-67	Sequence 67, Appl
27	27.8	18.5	615	US-09-328-111-32	Sequence 32, Appl

C	28	27.6	18.4	417	4	US-09-280-116-69	Sequence 69, Appl
	29	27.6	18.4	2963	3	US-09-232-200-60	Sequence 60, Appl
	30	27.6	18.4	2963	4	US-09-232-197-60	Sequence 60, Appl
	31	27.6	18.4	2963	4	US-09-232-201-60	Sequence 60, Appl
	32	27.6	18.4	2963	4	US-09-232-195-60	Sequence 60, Appl
	33	27.4	18.3	2241	4	US-09-134-001C-377	Sequence 377, App
	34	27.4	18.3	2411	3	US-09-188-930-75	Sequence 75, Appl
	35	27.4	18.3	2411	3	US-09-188-930-75	Sequence 75, Appl
	36	27.4	18.3	2411	4	US-09-312-283C-75	Sequence 256, Appl
	37	27.4	18.3	2411	4	US-09-312-283C-256	Sequence 256, App
	38	27.2	18.1	1837	2	US-08-933-750C-52	Sequence 52, Appl
	39	27.2	18.0	1837	3	US-09-234-613-52	Sequence 52, Appl
	40	27	18.0	531	4	US-09-621-976-10674	Sequence 10674, A
	C 41	26.8	17.9	591	4	US-09-702-705-1332	Sequence 1332, Ap
	C 42	26.8	17.9	591	4	US-09-702-705-1536	Sequence 1536, Ap
	C 43	26.8	17.9	591	4	US-09-736-457-1332	Sequence 1332, Ap
	C 44	26.8	17.9	591	4	US-09-736-457-1536	Sequence 1536, Ap
	C 45	26.8	17.9	591	4	US-09-614-124B-1332	Sequence 1332, Ap

ALIGNMENTS

```
RESULT 1
US-09-482-273-20/c
; Sequence 20, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; EARLIER FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 2525
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1354)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-482-273-20

Query Match 97.3%; Score 146; DB 4; Length 2525;
Best Local Similarity 95.3%; Pred. No. 2.8e-41;
Matches 143; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGATAGGAAACCCCTTAGGGCTGAGTGGAGCTGCGGACGACATATGCTTT 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1210 GAGATAGGAAACCCCTTAGGGCTGAGTGGAGCTGCGGACGACATATGCTTT 1151

QY 61 GTAAGACATGAGATGTTTATGTATGATATCTTAAAGACAGCACTTATCTTTAC 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1150 GTAAGACATGAGATGTTTATGTATGATATCTTAAAGACAGCACTTATCTTTAC 1091
```

QY 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
Db 1090 MTGTTATATGATGCAAGACCTTTGTTTCAC 1061

RESULT 2

US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Laufer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
; TITLE OF INVENTION: Sequences and Antibodies Thereo
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Flets, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HUA-H. CONTIG"
US-08-724-394A-20

Query Match 77.1%; Score 115.6; DB 2; Length 246240;

Best Local Similarity 90.0%; Pred. No. 9.9e-30; Mismatches 14; Indels 1; Gaps 1;

Matches 135; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 1 GAGATAGGGAAGAAACCGCTTAGAGGCTGAGGTGGGACCTGCGGGGACGAATATCTGCTT 60
Db 105006 GAGATAGGGAAGAAACCTGCTTAGAGGCTGAGGTGGGACGATGCTGGGACGAATATCTGCT 105065
QY 61 GTAAGACCTGAGATGTTTATGTTATGATGATATCTAAAGACAGCACTTAATCTTTTAC 120
Db 105066 TCAAGTCATGAGATGTTTATGTTATGATGATATCT-AAAGACAGCACTTAATCTTTTAC 105124

QY 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
Db 105125 CTGTTTATGATGCAAGACCTTTGTTTCAC 105154

RESULT 3

US-08-724-394A-21
; Sequence 21, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Laufer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
; TITLE OF INVENTION: Sequences and Antibodies Thereo
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Flets, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HUA-H. CONTIG"
US-08-724-394A-21

Query Match 77.1%; Score 115.6; DB 2; Length 246240;

Best Local Similarity 90.0%; Pred. No. 9.9e-30; Mismatches 14; Indels 1; Gaps 1;

Matches 135; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 1 GAGATAGGGAAGAAACCGCTTAGAGGCTGAGGTGGGACCTGCGGGGACGAATATCTGCTT 60
Db 105006 GAGATAGGGAAGAAACCTGCTTAGAGGCTGAGGTGGGACGATGCTGGGACGAATATCTGCT 105065
QY 61 GTAAGACCTGAGATGTTTATGTTATGATGATATCTAAAGACAGCACTTAATCTTTTAC 120
Db 105066 TCAAGTCATGAGATGTTTATGTTATGATGATATCT-AAAGACAGCACTTAATCTTTTAC 105124

QY 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
Db 105125 CTGTTTATGATGCAAGACCTTTGTTTCAC 105154

RESULT 4

US-08-724-394A-22
; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:

APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Laufer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Touchinashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereof
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Filts, Renee A. 35,136
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H-CONTIG"
US-08-724-394A-22
Query Match 77.1%; Score 115.6; DB 2; Length 246240;
Best Local Similarity 90.0%; Pred. No. 9.9e-30;
Matches 135; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
QY 1 GAGATAGGAAAAACCGCTTAGGGCTGAGAGTGGACCTCGGGCAGCAATACCTGCTT 60
DB 105006 GAGATAGGAAAAACCGCTTAGGGCTGAGAGTGGACCTCGGGCAGCAATACCTGCTT 105065
QY 61 GTAAGACCTGAGATGTTTATGTATGATGATCATCTAAAGACAGACACTTAATCCTTTAC 120
DB 105066 TCAAGCTATGAGATGTTTATGTATGATGATCATCT-AAAAGACAGCACTTAATCCTTTAC 105124
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 105125 CTGTGTTATGATGAGAGACCTTTGTTTAC 105154
RESULT 5
US-09-128-155-17
Sequence 17, Application US/09128155
Patent No. 6117654
GENERAL INFORMATION:
APPLICANT: Pan, Tang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128,155

CURRENT FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 176373
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(176373)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17
Query Match 71.2%; Score 106.8; DB 3; Length 176373;
Best Local Similarity 90.7%; Pred. No. 1e-26;
Matches 136; Conservative 0; Mismatches 12; Indels 2; Gaps 2;
QY 1 GAGATAGGAAAAACCGCTTAGGGCTGAGAGTGGACCTCGGGCAGCAATACCTGCTT 60
DB 175231 GAGATAGGAAAAACCGCTTAGGGCTGAGAGTGGACCTCGGGCAGCAATACCTGCTT 175289
QY 61 GTAAGACCTGAGATGTTTATGTATGATGATCATCTAAAGACAGACACTTAATCCTTTAC 120
DB 175290 TTAAGACATGAGATGTTTATGTATGATGATCATCT-AAAAGACAGCACTTAATCCTTTAC 175348
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 175349 CTGTGTTATGATGAGAGACCTTTGTTTAC 175378
RESULT 6
US-09-328-111-600
Sequence 600, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Dertl, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 600
LENGTH: 589
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(589)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-600
Query Match 61.9%; Score 92.8; DB 3; Length 589;
Best Local Similarity 80.7%; Pred. No. 6.1e-23;
Matches 121; Conservative 0; Mismatches 27; Indels 2; Gaps 1;

STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/598,326
FILING DATE: 20-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Poter, Jane E.R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 141:
SEQUENCE CHARACTERISTICS:
LENGTH: 938 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 141:
US-09-598-326-141

Query Match 50.7%; Score 76; DB 4; Length 938;
Best Local Similarity 78.0%; Pred. No. 1.6e-16;
Matches 117; Conservative 0; Mismatches 30; Indels 3; Gaps 2;
DB 1 GAGATAGGAGAAAACCCCTTAGGCTGAGGTGGAGCTCGGCGACGCAATACCTGCTT 60
DB 608 GAGATAGGAGAAAACCTTAGGCTGAGGTGGAGCTCGGCGACGCAATACCTGCTT 667
DB 61 GTAAGACCTGAGATGTTTATGTATGATGATATCTAAAGACGACCTTAATCTTTAC 120
DB 668 TTAATGACCGAGATGTTTATGATGACATC--AAGCAGACGACCT-TTCCTTAAA 724
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 725 CTTATTATGACACAGACCTTTGTTTAC 754

RESULT 10
US-09-289-198-141
Sequence 141, Application US/09289198
Patent No. 6586570
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
APPLICANT: Smith, John M.
APPLICANT: Reed, Steven G.
APPLICANT: Misher, Lynda
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
FILE REFERENCE: 210121.419C5
CURRENT APPLICATION NUMBER: US/09/289,198
CURRENT FILING DATE: 1999-04-09
EARLIER APPLICATION NUMBER: US 09/062,451
EARLIER FILING DATE: 1998-04-17
EARLIER APPLICATION NUMBER: US 08/991,789
EARLIER FILING DATE: 1997-12-11
EARLIER APPLICATION NUMBER: US 08/838,762
EARLIER FILING DATE: 1997-04-09
EARLIER APPLICATION NUMBER: PCT/US97/00485
EARLIER FILING DATE: 1997-01-10
EARLIER APPLICATION NUMBER: US 08/700,014
EARLIER FILING DATE: 1996-08-20
EARLIER APPLICATION NUMBER: US 08/585,392
EARLIER FILING DATE: 1996-01-01

NUMBER OF SEQ ID NOS: 312
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 141
LENGTH: 938
TYPE: DNA
ORGANISM: Homo sapien
US-09-289-198-141

Query Match 50.7%; Score 76; DB 4; Length 938;
Best Local Similarity 78.0%; Pred. No. 1.6e-16;
Matches 117; Conservative 0; Mismatches 30; Indels 3; Gaps 2;
DB 1 GAGATAGGAGAAAACCCCTTAGGCTGAGGTGGAGCTCGGCGACGCAATACCTGCTT 60
DB 608 GAGATAGGAGAAAACCTTAGGCTGAGGTGGAGCTCGGCGACGCAATACCTGCTT 667
DB 61 GTAAGACCTGAGATGTTTATGTATGATGATATCTAAAGACGACCTTAATCTTTAC 120
DB 668 TTAATGACCGAGATGTTTATGATGACATC--AAGCAGACGACCT-TTCCTTAAA 724
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 725 CTTATTATGACACAGACCTTTGTTTAC 754

RESULT 11
US-09-429-755-141
Sequence 141, Application US/09429755A
Patent No. 6656480
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
APPLICANT: Smith, John M.
APPLICANT: Reed, Steven G.
APPLICANT: Misher, Lynda
APPLICANT: Retter, Marc W.
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
FILE REFERENCE: 210121.419C6
CURRENT APPLICATION NUMBER: US/09/429,755A
CURRENT FILING DATE: 1999-10-28
NUMBER OF SEQ ID NOS: 315
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 141
LENGTH: 938
TYPE: DNA
ORGANISM: Homo sapien
US-09-429-755-141

Query Match 50.7%; Score 76; DB 4; Length 938;
Best Local Similarity 78.0%; Pred. No. 1.6e-16;
Matches 117; Conservative 0; Mismatches 30; Indels 3; Gaps 2;
DB 1 GAGATAGGAGAAAACCCCTTAGGCTGAGGTGGAGCTCGGCGACGCAATACCTGCTT 60
DB 608 GAGATAGGAGAAAACCTTAGGCTGAGGTGGAGCTCGGCGACGCAATACCTGCTT 667
DB 61 GTAAGACCTGAGATGTTTATGTATGATGATATCTAAAGACGACCTTAATCTTTAC 120
DB 668 TTAATGACCGAGATGTTTATGATGACATC--AAGCAGACGACCT-TTCCTTAAA 724
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 725 CTTATTATGACACAGACCTTTGTTTAC 754

RESULT 12
US-09-904-615-45/C
Sequence 45, Application US/09904615
Patent No. 6566325
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 49 Human Secreted Proteins

FILE REFERENCE: P2032P1
CURRENT APPLICATION NUMBER: US/09/904,615
CURRENT FILING DATE: 2001-07-16
PRIOR APPLICATION NUMBER: 09/511,554
PRIOR FILING DATE: 2000-02-23
PRIOR APPLICATION NUMBER: 60/097,917
PRIOR FILING DATE: 1998-08-25
PRIOR APPLICATION NUMBER: 60/098,634
PRIOR FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 170
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO: 45
LENGTH: 2622
TYPE: DNA
ORGANISM: Homo sapiens
US-09-904-615-45

Query Match 34.4%; Score 51.6; DB 4; Length 2622;
Best Local Similarity 67.9%; Pred. No. 3.2e-08;
Matches 72; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAACCGCTTAGGCTGAGAGTGAGCGGAGCAGCAATACCTGCTT 60
DB 1358 GAGATAGGAAAAACCGCTTAGGCTGAGAGTGAGCGGAGCAGCAATACCTGCTT 1299
QY 61 GTAAAGCACTGAGATGTTTATGTGTATGATCATATCTTAAAGCAGAC 106
DB 1298 GTTGCTCTGCTAGTAGATATTTGTGTAAAGTAAACATTAATC 1253

RESULT 13
US-08-991-789A-240
Sequence 240, Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:

APPLICANT: Prudakis, Tony N.
Smith, John M.

REED, Steven G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER

NUMBER OF SEQUENCES: 292

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed IP Law Group

STREET: 701 Fifth Avenue, Suite 6300

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/991,789A

FILING DATE: 11-Dec-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Potter, Jane E. R.

REGISTRATION NUMBER: 33,332

REFERENCE/DOCKET NUMBER: 210121.419C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 240:

SEQUENCE CHARACTERISTICS:

LENGTH: 519 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: cDNA

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

SEQUENCE DESCRIPTION: SEQ ID NO: 240:
US-08-991-789A-240

Query Match 26.1%; Score 39.2; DB 3; Length 519;
Best Local Similarity 58.6%; Pred. No. 0.00033;
Matches 68; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 7 GGGAAAAACCGCTTAGGCTGAGAGTGAGCGGAGCAGCAATACCTTTGTAAG 66
DB 227 GAGAAAAACCGCTTAGGCTGAGAGTGAGCGGAGCAGCAATACCTTTGTAAG 286
QY 67 CACTGAGATTTATGTATGATATCTTAAAGCAGACACTTAATCCTTACAT 122
DB 287 CTTTACTCCACAGATTTTGGCGGAGGAGAAACATTAATCTGGCTTACGTCACAT 342

RESULT 14
US-09-062-451-240
Sequence 240, Application US/09062451
Patent No. 6344550
GENERAL INFORMATION:

APPLICANT: Prudakis, Tony N.

SMITH, John M.

REED, Steven G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER

NUMBER OF SEQUENCES: 297

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/062,451

FILING DATE: 04-APR-1997

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.419C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

INFORMATION FOR SEQ ID NO: 240:

SEQUENCE CHARACTERISTICS:

LENGTH: 519 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: cDNA

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

Query Match 26.1%; Score 39.2; DB 4; Length 519;
Best Local Similarity 58.6%; Pred. No. 0.00033;
Matches 68; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 7 GGGAAAAACCGCTTAGGCTGAGAGTGAGCGGAGCAGCAATACCTTTGTAAG 66
DB 227 GAGAAAAACCGCTTAGGCTGAGAGTGAGCGGAGCAGCAATACCTTTGTAAG 286
QY 67 CACTGAGATTTATGTATGATATCTTAAAGCAGACACTTAATCCTTACAT 122
DB 287 CTTTACTCCACAGATTTTGGCGGAGGAGAAACATTAATCTGGCTTACGTCACAT 342

This Page Blank (Aspirin)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2004, 22:35:13 ; Search time 65.3061 Seconds
(without alignments)
8285.900 Million cell updates/sec

Title: US-10-016-604-155

Perfect score: 150
Sequence: 1 gagatagggaacacgcgcct.....atgcgaagacctgtcttcac 150

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2353733 seqs, 180373377 residues

Total number of hits satisfying chosen parameters: 4707466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148.4	98.9	35414	9	US-09-836-607-45
2	148.4	98.9	35414	10	US-09-421-112-82
3	146.8	97.9	32249	10	US-09-764-891-8024
4	146	97.3	2525	10	US-09-984-271-20
5	145.2	96.8	390	15	US-10-242-535A-15659
6	143.6	95.7	744802	15	US-10-292-798-1369
7	142	94.7	2398	15	US-10-027-632-101981
8	142	94.7	2398	15	US-10-027-632-101982
9	142	94.7	2398	15	US-10-027-632-101983
10	140.4	93.6	975	15	US-10-233-958-17
11	140.4	93.6	2689	15	US-10-233-958-4
12	140.4	93.6	6020	15	US-10-233-958-26
13	140.4	93.6	10569	15	US-10-233-958-42
14	140.4	93.6	1691139	14	US-10-067-514-1
15	140.4	93.6	1691139	15	US-10-419-723-1

C	16	138.8	92.5	619	15	US-10-027-632-138010	Sequence 138010,
	17	138.8	92.5	969	15	US-10-233-958-18	Sequence 18, Appl
	18	138.8	92.5	972	15	US-10-233-958-20	Sequence 20, Appl
	19	138.8	92.5	1010	15	US-10-233-958-19	Sequence 19, Appl
	20	138.8	92.5	3112	15	US-10-027-632-114936	Sequence 114936,
	21	137.2	91.5	975	15	US-10-233-958-15	Sequence 15, Appl
	22	137.2	91.5	975	15	US-10-233-958-16	Sequence 16, Appl
	23	137.2	91.5	1148	15	US-09-764-891-6785	Sequence 6785, Ap
	24	137.2	91.5	1148	14	US-10-091-572-632	Sequence 632, Ap
	25	137.2	91.5	2689	15	US-10-233-958-2	Sequence 2, Appl
	26	137.2	91.5	9343	15	US-10-233-958-43	Sequence 43, Appl
	27	134.6	89.7	1574	15	US-10-242-355-1080	Sequence 1080, Ap
	28	132.4	88.3	2189	15	US-10-027-632-101674	Sequence 101674,
	29	132.6	81.7	1043	9	US-09-764-869-1485	Sequence 1485, Ap
	30	132.6	81.7	1043	14	US-10-091-504-1485	Sequence 1485, Ap
	31	122.6	81.7	1043	15	US-10-227-577-1485	Sequence 1485, Ap
	32	122	81.3	2495	5	US-10-027-632-103060	Sequence 103060,
	33	122	81.3	198285	9	US-09-880-107-3814	Sequence 3814, Ap
	34	121	80.7	922	9	US-09-764-869-1486	Sequence 1486, Ap
	35	121	80.7	922	14	US-10-091-504-1486	Sequence 1486, Ap
	36	121	80.7	922	15	US-10-227-577-1486	Sequence 1486, Ap
	37	117.8	78.5	2126	15	US-10-027-632-98430	Sequence 98430, A
	38	117.2	78.1	855	10	US-09-764-891-8343	Sequence 8343, Ap
	39	117.2	78.1	855	14	US-10-205-428-776	Sequence 776, Ap
	40	115.6	77.1	855	10	US-09-764-891-8342	Sequence 8342, Ap
	41	115.6	77.1	855	14	US-10-205-428-775	Sequence 775, Ap
	42	115.6	77.1	17588	14	US-10-017-161-1603	Sequence 1603, Ap
	43	115.6	77.1	17588	15	US-10-292-798-1277	Sequence 1277, Ap
	44	115.6	77.1	235033	14	US-10-301-844-1	Sequence 1, Appl
	45	115.6	77.1	237326	14	US-10-301-844-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-836-607-45
Sequence 45, Application US/09836607
Patent No. US2002098541A1
GENERAL INFORMATION:
APPLICANT: NI, Jian
TITLE OF INVENTION: TMR Related Gene 12
FILE REFERENCE: P4990P1
CURRENT APPLICATION NUMBER: US/09/836,607
CURRENT FILING DATE: 2001-04-17
PRIOR APPLICATION NUMBER: 2000-04-19
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: 09/421,112
PRIOR FILING DATE: 1999-10-19
PRIOR APPLICATION NUMBER: 60/104,950
PRIOR FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 47
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 45
LENGTH: 35414
TYPE: DNA
ORGANISM: Homo sapiens
US-09-836-607-45

Query Match 98.9%; Score 148.4; DB 9; Length 35414;
Best Local Similarity 99.3%; Pred. No. 7.9e-39;
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GAGATAGGAAAAACCCCTTAGGCTGAGAGTGGAGCTTGGGAGCAATATCGCTTT 60
DB 9241 GAGTATGAGAAAAACCCCTTAGGCTGAGAGTGGAGCTTGGGAGCAATATCGCTTT 9300
OY 61 GTAAGACATGAGATGTTATGTGTATGATGATCTAAAGACAGACTTAATCTTTAC 120
DB 9301 GTAAGACATGAGATGTTATGTGTATGATGATCTAAAGACAGACTTAATCTTTAC 9360
OY 121 ATTGTATATGATGCAAGACCTTTGTTTCAC 150

Db 9361 ATTGCTATGATGCAGAACCTTTGTTTCAC 9390

RESULT 2

US-09-421-112-45
; Sequence 45, Application US/09421112
; Publication No. US20030082532A1
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: TNFR Related Gene 12
; FILE REFERENCE: P490
; CURRENT APPLICATION NUMBER: US/09/421,112
; CURRENT FILING DATE: 1999-10-19
; EARLIER APPLICATION NUMBER: 60/104,950
; EARLIER FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 35414
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-421-112-45

Query Match 98.9%; Score 148.4; DB 10; Length 35414;
Best Local Similarity 99.3%; Pred. No. 7.9e-39;
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGATAGGAAAAACCGCTTAGGGCTGAGAGTGGACCTGGGGCAGCAATACCTGCTTT 60

Db 9241 GAGATAGGAAAAACCGCTTAGGGCTGAGAGTGGACCTGGGGCAGCAATACCTGCTTT 9300

Qy 61 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120

Db 9301 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 9360

Qy 121 ATTGCTATGATGCAGAACCTTTGTTTCAC 150

Db 9361 ATTGCTATGATGCAGAACCTTTGTTTCAC 9390

RESULT 3

US-09-764-891-8024
; Sequence 8024, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8024
; LENGTH: 32249
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-8024

Query Match 97.9%; Score 146.8; DB 10; Length 32249;
Best Local Similarity 98.7%; Pred. No. 2.6e-38;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAGATAGGAAAAACCGCTTAGGGCTGAGAGTGGACCTGGGGCAGCAATACCTGCTTT 60

Db 26531 GAGATAGGAAAAACCGCTTAGGGCTGAGAGTGGACCTGGGGCAGCAATACCTGCTTT 26590

Qy 61 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120

Db 26591 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 26650

Qy 121 ATTGCTATGATGCAGAACCTTTGTTTCAC 150

Db 26651 ATTGCTATGATGCAGAACCTTTGTTTCAC 26680

RESULT 4

US-09-984-271-20/c
; Sequence 20, Application US/09984271
; Publication No. US20030040088A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/984,271
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/482,273
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US99/15849
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: 60/092,921
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,922
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,956
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 2525
; TYPE: DNA
; ORGANISM: Homo sapiens

FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1354)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-984-271-20

Query Match 97.3%; Score 146; DB 10; Length 2525;
Best Local Similarity 95.3%; Pred. No. 1.4e-38;
Matches 143; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGATAGGAAAAACCGCTTAGGGCTGAGAGTGGACCTGGGGCAGCAATACCTGCTTT 60

Db 1210 GAGATAGGAAAAACCGCTTAGGGCTGAGAGTGGACCTGGGGCAGCAATACCTGCTTT 1151

Qy 61 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120

Db 1150 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 1091

Qy 121 ATTGCTATGATGCAGAACCTTTGTTTCAC 150

Db 1090 MTTGTYTATGATGCAGAACCTTTGTTTCAC 1061

RESULT 5

US-10-242-535A-15659
; Sequence 15659, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28

```

; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15659
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Human
; US-10-242-535A-15659

Query Match          96.8%; Score 145.2; DB 15; Length 390;
Best Local Similarity 98.0%; Pred. No. 1e-38;
Matches 147; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAACCCGCTTAGGGCTGAGAGTGAGACCTGCGGCGAGCAATACCTGCTT 60
   |||||
DB 75 GAGATTAGGAAAAACCCGCTTAGGGCTGAGAGTGAGACCTGCGGCGAGCAATACCTGCTT 134
   |||||

QY 61 GTAAGACCTGAGATGTTATGTATGATGATATCTAAAGACAGACCTTAATCCTTTAC 120
   |||||
DB 135 GTAAGACCTGAGATGTTATGTATGATGATATCTAAAGACAGACCTTAATCCTTTAC 194
   |||||

QY 121 ATTGCTATATGATGCAAGACCTTGTTCAC 150
   |||||
DB 195 ATTGCTATATGATGCAAGACCTTGTTCAC 224
   |||||

RESULT 6
US-10-292-798-1369/C
; Sequence 1369, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUMA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABRARANT, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1369
; LENGTH: 744802
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(744802)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(246)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (25640)..(25677)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27078)..(27094)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (141192)..(141769)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159571)..(159606)
; LOCATION: (159571)..(159606)
```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (174525)..(174575)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (234891)..(235013)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (235514)..(235560)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (279677)..(279729)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (408660)..(409123)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (409204)..(409669)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (428381)..(428396)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (472204)..(472330)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (714252)..(714355)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (714447)..(714529)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (739794)..(739891)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (744484)..(744602)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (51812)..(51911)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (57122)..(57221)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (79368)..(79467)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (293951)..(294050)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (31089)..(310188)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (332935)..(332935)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (332992)..(332992)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (362002)..(362101)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (639781)..(639880)
; OTHER INFORMATION: a, t, c, g, unknown or other
; US-10-292-798-1369
```

Query Match 95.7%; Score 143.6; DB 15; Length 744802;
Best Local Similarity 97.3%; Pred. No. 1.5e-36;
Matches 146; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAACCCCTTAGGCTGAGGTGGGACCTCGGGGACGCAATACCTGCTT 60
DB 635834 GAGATAGGAAAAACCCCTTAGGCTGAGGTGGGACGCAATACCTGCTT 635775
QY 61 GTAAGACGACGAGTGTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120
DB 635774 GTAAGACGAGTGTATGTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 635715
QY 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
DB 635714 CTGTCTATGATGCAAGACCTTTGTTTCAC 635685

RESULT 7

US-10-027-632-101981
; Sequence 101981, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101981
; LENGTH: 2398
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-101981

Query Match 94.7%; Score 142; DB 15; Length 2398;
Best Local Similarity 96.7%; Pred. No. 3e-37; 5; Indels 0; Gaps 0;
Matches 145; Conservative 0; Mismatches 5;

QY 1 GAGATAGGAAAAACCCCTTAGGCTGAGGTGGGACCTCGGGGACGCAATACCTGCTT 60
DB 2136 GAGATAGGAAAAACCCCTTAGGCTGAGGTGGGACGCAATACCTGCTT 2195
QY 61 GTAAGACGACGAGTGTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120
DB 2196 GTAAGACGAGTGTATGTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 2255
QY 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
DB 2256 CTGTCTATGATGCAAGACCTTTGTTTCAC 2285

RESULT 8

US-10-027-632-101982
; Sequence 101982, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101982
; LENGTH: 2398
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-101982

Query Match 94.7%; Score 142; DB 15; Length 2398;
Best Local Similarity 96.7%; Pred. No. 3e-37; 5; Indels 0; Gaps 0;
Matches 145; Conservative 0; Mismatches 5;

QY 1 GAGATAGGAAAAACCCCTTAGGCTGAGGTGGGACCTCGGGGACGCAATACCTGCTT 60
DB 2136 GAGATAGGAAAAACCCCTTAGGCTGAGGTGGGACGCAATACCTGCTT 2195
QY 61 GTAAGACGACGAGTGTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120
DB 2196 GTAAGACGAGTGTATGTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 2255
QY 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
DB 2256 CTGTCTATGATGCAAGACCTTTGTTTCAC 2285

RESULT 9

US-10-027-632-101983
; Sequence 101983, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101983

LENGTH: 2398
TYPE: DNA
ORGANISM: Human
US-10-027-632-101983

Query Match 94.7%; Score 142; DB 15; Length 2398;
Best Local Similarity 96.7%; Pred. No. 3e-37;
Matches 145; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 60
Db 2136 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 2195
Qy 61 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120
Db 2196 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 2255
Qy 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
Db 2256 CTGTCTATGATGCAAGACCTTTGTTTAC 2285

RESULT 10
US-10-233-958-17
Sequence 17, Application US/10233958
Publication No. US20040009468A1
GENERAL INFORMATION:
APPLICANT: Mach, Bernard
APPLICANT: Conrad, Bernard
TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
TITLE OF INVENTION: Thereof and Use in the Determination of Genetic
TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
TITLE OF INVENTION: Provirus
FILE REFERENCE: 23135-504
CURRENT APPLICATION NUMBER: US/10/233, 958
CURRENT FILING DATE: 2002-09-03
PRIOR APPLICATION NUMBER: 60/316, 513
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: 60/316, 522
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 17
LENGTH: 975
TYPE: DNA
ORGANISM: Human endogenous retrovirus
US-10-233-958-17

Query Match 93.6%; Score 140.4; DB 15; Length 975;
Best Local Similarity 96.0%; Pred. No. 6.6e-37;
Matches 144; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 60
Db 562 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 621
Qy 61 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120
Db 622 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 681
Qy 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
Db 682 CTGTCTATGATGCAAGACCTTTGTTTAC 711

RESULT 11
US-10-233-958-4
Sequence 4, Application US/10233958
Publication No. US20040009468A1
GENERAL INFORMATION:
APPLICANT: Mach, Bernard
APPLICANT: Conrad, Bernard
TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis

TITLE OF INVENTION: Thereof and Use in the Determination of Genetic
TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
TITLE OF INVENTION: Provirus
FILE REFERENCE: 23135-504
CURRENT APPLICATION NUMBER: US/10/233, 958
CURRENT FILING DATE: 2002-09-03
PRIOR APPLICATION NUMBER: 60/316, 513
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: 60/316, 522
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 2689
TYPE: DNA
ORGANISM: Human endogenous retrovirus
US-10-233-958-4

Query Match 93.6%; Score 140.4; DB 15; Length 2689;
Best Local Similarity 96.0%; Pred. No. 1.1e-36;
Matches 144; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 60
Db 2282 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 2341
Qy 61 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120
Db 2342 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 2401
Qy 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
Db 2402 CTGTCTATGATGCAAGACCTTTGTTTAC 2431

RESULT 12
US-10-233-958-26/c
Sequence 26, Application US/10233958
Publication No. US20040009468A1
GENERAL INFORMATION:
APPLICANT: Mach, Bernard
APPLICANT: Conrad, Bernard
TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
TITLE OF INVENTION: Thereof and Use in the Determination of Genetic
TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
TITLE OF INVENTION: Provirus
FILE REFERENCE: 23135-504
CURRENT APPLICATION NUMBER: US/10/233, 958
CURRENT FILING DATE: 2002-09-03
PRIOR APPLICATION NUMBER: 60/316, 513
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: 60/316, 522
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 26
LENGTH: 6020
TYPE: DNA
ORGANISM: Human endogenous retrovirus
US-10-233-958-26

Query Match 93.6%; Score 140.4; DB 15; Length 6020;
Best Local Similarity 96.0%; Pred. No. 1.6e-36;
Matches 144; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 60
Db 2194 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 2135
Qy 61 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120
Db 2134 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 2075

Qy 121 ATTGCTATGATGCAAGACCTTGTTCAC 150
Db 2074 CTGTCTATGATGCAAGACCTTGTTCAC 2045

RESULT 13
US-10-233-958-42
Sequence 42. Application US/10233958

```

Yy      121  ATTGCTATGATGCAAGACCTTTGTTAC 150
Db      2074  CTTCCTATGATGCAAGACCTTTGTTAC 2045

RESULT 13
US-10-233-958-42
/ Sequence 42, Application US/10233958
/ Publication No. US20040009468A1
/ GENERAL INFORMATION:
/ APPLICANT: Mach, Bernard
/ APPLICANT: Conrad, Bernard
/ TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
/ TITLE OF INVENTION: thereof and Use in the Determination of Genetic
/ TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
/ FILE REFERENCE: 23135-504
/ CURRENT APPLICATION NUMBER: US/10/233,958
/ CURRENT FILING DATE: 2002-09-03
/ PRIOR APPLICATION NUMBER: 60/316,513
/ PRIOR FILING DATE: 2001-08-31
/ PRIOR APPLICATION NUMBER: 60/316,522
/ PRIOR FILING DATE: 2001-08-31
/ NUMBER OF SEQ ID NOS: 46
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 42
/ LENGTH: 10569
/ TYPE: DNA
/ ORGANISM: Human endogenous retrovirus
/ US-10-233-958-42

```

Result No.	Score	Query Match	Length	DB	ID	Description
1	856.6	97.5	32249	10	US-09-764-891-8024	Sequence 8024, App1
2	845.4	96.2	35414	9	US-09-836-607-45	Sequence 45, App1
3	845.4	96.2	35414	10	US-09-421-112-45	Sequence 45, App1
4	805.6	91.6	1691139	14	US-10-067-514-1	Sequence 1, App1
5	805.6	91.6	1691139	15	US-10-419-723-1	Sequence 1, App1
6	797.6	90.7	975	15	US-10-233-958-15	Sequence 15, App1
7	797.6	90.7	975	15	US-10-233-958-17	Sequence 17, App1
8	797.6	90.7	2689	15	US-10-233-958-2	Sequence 2, App1
9	797.6	90.7	2689	15	US-10-233-958-4	Sequence 4, App1
10	796	90.6	972	15	US-10-233-958-20	Sequence 20, App1
11	796	90.6	6030	15	US-10-233-958-6	Sequence 26, App1
12	796	90.6	10569	15	US-10-233-958-8	Sequence 42, App1
13	794.4	90.4	969	15	US-10-233-958-18	Sequence 18, App1
14	794.4	90.4	975	15	US-10-233-958-16	Sequence 16, App1
15	792.8	90.2	1010	15	US-10-233-958-19	Sequence 19, App1

C	16	789.2	89.8	744502	15	US-10-232-798-1369	Sequence 1369, Ap
C	17	789.4	89.2	198285	9	US-09-880-107-3814	Sequence 3814, Ap
C	18	788.2	89.0	9343	15	US-10-233-958-43	Sequence 43, Appl
C	19	778.6	88.6	1574	15	US-10-242-355-1080	Sequence 1080, Ap
C	20	764.8	87.0	1043	9	US-09-764-868-1485	Sequence 1485, Ap
C	21	764.8	87.0	1043	14	US-10-091-504-1485	Sequence 1485, Ap
C	22	764.8	87.0	1043	15	US-10-227-577-1485	Sequence 1485, Ap
C	23	763.8	86.9	2189	15	US-10-027-632-101674	Sequence 101674,
C	24	763.8	86.8	922	9	US-09-764-868-1486	Sequence 1486, Ap
C	25	763.2	86.8	922	14	US-10-091-504-1486	Sequence 1486, Ap
C	26	763.2	86.8	922	15	US-10-227-577-1486	Sequence 1486, Ap
C	27	748.8	85.2	54433	15	US-10-085-117-124	Sequence 124, Appl
C	28	738	84.0	2398	15	US-10-027-632-101981	Sequence 101981,
C	29	738	84.0	2398	15	US-10-027-632-101981	Sequence 101982,
C	30	738	84.0	2398	15	US-10-027-632-101983	Sequence 101983,
C	31	735.2	83.6	17633	13	US-10-095-407-17	Sequence 17, Appl
C	32	721.2	82.0	3112	15	US-10-027-632-114936	Sequence 114936,
C	33	719.8	81.9	1148	10	US-09-764-891-6785	Sequence 6785, Ap
C	34	719.8	81.9	1148	14	US-10-091-572-632	Sequence 632, Appl
C	35	717.4	81.6	235033	14	US-10-301-844-1	Sequence 1, Appl1
C	36	717.4	81.6	237326	14	US-10-301-844-2	Sequence 2, Appl1
C	37	696	79.2	2126	15	US-10-027-632-98430	Sequence 98430, Appl
C	38	692.2	78.7	2525	10	US-09-984-271-20	Sequence 20, Appl
C	39	688.6	78.3	2454	15	US-10-04-047-275	Sequence 275, Appl
C	40	682	77.6	855	10	US-09-764-891-8343	Sequence 8343, Ap
C	41	682	77.6	855	14	US-10-205-428-776	Sequence 776, Appl
C	42	680.4	77.4	855	10	US-09-764-891-8342	Sequence 8342, Ap
C	43	680.4	77.4	855	14	US-10-205-428-775	Sequence 775, Appl
C	44	608.8	69.3	1039	15	US-10-027-632-119427	Sequence 119427,
C	45	608.8	69.3	1039	15	US-10-027-632-119428	Sequence 119428,

ALIGNMENTS

```

RESULT 1
US-09-764-891-8024
; Sequence 8024, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; PRIORITY FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8024
; LENGTH: 32249
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-8024

```

Query Match 97.5%; Score 856.6; DB 10; Length 32249;

Matches 865; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy	1	TGTGGGGAAAAGCAAGAGATCAAGATTGTTACTGTGTCTGTGTATGAAAAGAAAGTAAACAT	60
Db	25971	TGTGGGGAAAAGCAAGAGATCAAGATTGTTACTGTGTCTGTGTATGAAAAGAAAGTAAACAT	26033
Qy	61	AGGAGACTCCATTGTTGTATGTAATAAGAAAATTTCTGCCTTGAGATTCTGTTAATC	120
Db	26031	AGGAGACTCCATTGTTGTATGTAATAAGAAAATTTCTGCCTTGAGATTCTGTTAATC	26099
Qy	121	TATGACCTTAACCCCAACCCCGTCTCTGTAACATGTGCTGTGCACACTCAGGGTTAA	180
Db	26091	TATTAACCTTAACCCCAACCCCGTCTCTGTAACATGTGCTGTGCACACTCAGAGTTGA	26159
Qy	181	ATGATTAATAGGGGGGTGAGAGATGTCCTTGTTTAAACAATGTTTAAGACGACATGTC	240
Db	26151	ATGATTAATAGGGGGGTGAGAGATGTCCTTGTTTAAACAATGTTTAAGACGACATGTC	26210

QY 241 CTTAAGATCATCACCACTCCCTAATCTCAAGTACCCAGGAGACAAAACCTGCGAGG 300
DB 26211 CTTAAGATCATCACCACTCCCTAATCTCAAGTACCCAGGAGACAAAACCTGCGAGG 26270
QY 301 CCGCAGGAGCTCTGCTCTAGAGAAAGCAGGATTTGTCACGTTTCTCCCATGTGTAG 360
DB 26271 CCGCAGGAGCTCTGCTCTAGAGAAAGCAGGATTTGTCACGTTTCTCCCATGTGTAG 26330
QY 361 CCGAATATAGGCTGCTGAGGAGGAGAACCGGACCGTCCCGACCGGACACCGGTA 420
DB 26331 TCTGAATATAGGCTGCTGAGGAGGAGAACCGGACCGTCCCGACCGGACACCGGTA 26390
QY 421 AAGGCTCTGCTGAGGAGGATTAAGTAAAGAGAGAAATGCTCTTGGAGTTGAGACA 480
DB 26391 AAGGCTCTGCTGAGGAGGATTAAGTAAAGAGAGAAATGCTCTTGGAGTTGAGACA 26450
QY 481 AGAGAAAGCATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 26451 AGAGAAAGCATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26510
QY 541 GATTGATGCTCATCTGAGTGAAGTGAAGGAGAAACCGGCTTGAAGGCTGAGGAGCT 600
DB 26511 GATTGATGCTCATCTGAGTGAAGTGAAGGAGAAACCGGCTTGAAGGCTGAGGAGCT 26570
QY 601 GCGGAGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
DB 26571 GCGGAGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26630
QY 661 CACAGCACTTAATCTTATCAATGCTGATGATGCAAGACCTTGTTCACATGTTGCT 720
DB 26631 CACAGCACTTAATCTTATCAATGCTGATGATGCAAGACCTTGTTCACATGTTGCT 26690
QY 721 GCGACCTCTCCCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 26691 GCGACCTCTCCCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26750
QY 781 CCACAGATGATCAATAATCTAAGGAACTCAGAGGCTGCGGAGTCTCCATATGCTG 840
DB 26751 CCACAGATGATCAATAATCTAAGGAACTCAGAGGCTGCGGAGTCTCCATATGCTG 26810
QY 841 AACGCTGTTCCCGGGTCCCTTCTTCTTCTCTATA 879
DB 26811 AACGCTGTTCCCGGGTCCCTTCTTCTTCTCTATA 26849

RESULT 2
US-09-836-607-45
; Sequence 45, Application US/09836607
; Patent No. US20020098541A1
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: TNFR Related Gene 12
; FILE REFERENCE: PF490P1
; CURRENT APPLICATION NUMBER: US/09/836, 607
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 60/198,388
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 09/421,112
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/104,950
; PRIOR FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 45
; LENGTH: 35414
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-836-607-45

Query Match 96.2%; Score 845.4; DB 9; Length 35414;
Best Local Similarity 97.6%; Pred. No. 1,26-274;

Matches 858; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 1 TGTGGGAAAGCAAGAGATGATGATTTGATCTGTCTGTGTAGAAAGATAGCAT 60
DB 8681 TGTGGGAAAGCAAGAGATGATGATTTGATCTGTCTGTGTAGAAAGATAGCAT 8740
QY 61 AGGAGATCTCCATTTGTATGTAATAAGAAATTTCTTCTGCTTGAAGATTTCTTTATC 120
DB 8741 AGGAGATCTCCATTTGTATGTAATAAGAAATTTCTTCTGCTTGAAGATTTCTTTATC 8800
QY 121 TATGACCTTACCCCAACCCCGGCTCTGAAACATNGCTGTCATCAAGGTTAA 180
DB 8801 TATGACCTTACCCCAACCCCGGCTCTGAAACATNGCTGTCATCAAGGTTAA 8860
QY 181 ATGATTAAGGGGCGGTGAGATGCTCTTGTAAACAGATGCTTGAAGCAGATGCTC 240
DB 8861 ATGATTAAGGGGCGGTGAGATGCTCTTGTAAACAGATGCTTGAAGCAGATGCTC 8920
QY 241 CTTAAGATCATCACCACTCCCTAATCTCAAGTACCCAGGAGACAAAACCTGCGAAG 300
DB 8921 CTTAAGATCATCACCACTCCCTAATCTCAAGTACCCAGGAGACAAAACCTGCGAAG 8980
QY 301 CCGCAGGAGCTCTGCTCTAGAGAAAGCAGGATTTGTCACGTTTCTCCCATGTGTAG 360
DB 8981 CCGCAGGAGCTCTGCTCTAGAGAAAGCAGGATTTGTCACGTTTCTCCCATGTGTAG 9040
QY 361 CCGAATATAGGCTGCTGAGGAGGAGAAAGACCTGACCGTCCCGACCGGACCGGTA 420
DB 9041 CCGAATATAGGCTGCTGAGGAGGAGAAAGACCTGACCGTCCCGACCGGACCGGTA 9100
QY 421 AAGGCTCTGCTGAGGAGGATTAAGTAAAGAGAGAAATGCTCTTGCAGTTGAGACA 480
DB 9101 AAGGCTCTGCTGAGGAGGATTAAGTAAAGAGAGAAATGCTCTTGCAGTTGAGACA 9160
QY 481 AGAGAAAGCATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 9161 AGAGAAAGCATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9220
QY 541 GATTGATGCTCATCTGAGTGAAGTGAAGGAGAAACCGGCTTGAAGGCTGAGGAGCT 600
DB 9221 GATTGATGCTCATCTGAGTGAAGTGAAGGAGAAACCGGCTTGAAGGCTGAGGAGCT 9280
QY 601 GCGGAGCAATCTGCTTGTAAAGCACTGAGATGTTATGCTATGATCAATCTAAAG 660
DB 9281 GCGGAGCAATCTGCTTGTAAAGCACTGAGATGTTATGCTATGATCAATCTAAAG 9340
QY 661 CACAGCACTTAATCTTATCAATGCTGATGATGCAAGACCTTGTTCACATGTTGCT 720
DB 9341 CACAGCACTTAATCTTATCAATGCTGATGATGCAAGACCTTGTTCACATGTTGCT 9400
QY 721 GCTGACCTCTCCCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 9401 GCTGACCTCTCCCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9460
QY 781 CCACAGATGATCAATAATCTAAGGAACTCAGAGGCTGCGGAGTCTCCATATGCTG 840
DB 9461 CCACAGATGATCAATAATCTAAGGAACTCAGAGGCTGCGGAGTCTCCATATGCTG 9520
QY 841 AACGCTGTTCCCGGGTCCCTTCTTCTTCTCTATA 879
DB 9521 AACGCTGTTCCCGGGTCCCTTCTTCTTCTCTATA 9559

RESULT 3
US-09-421-112-45
; Sequence 45, Application US/09421112
; Publication No. US20030082532A1
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: TNFR Related Gene 12
; FILE REFERENCE: PF490
; CURRENT APPLICATION NUMBER: US/09/421,112

;; CURRENT FILING DATE: 1999-10-19
;; EARLIER APPLICATION NUMBER: 60/104,950
;; EARLIER FILING DATE: 1998-10-20
;; NUMBER OF SEQ ID NOS: 47
;; SOFTWARE: Patent Ver. 2.0
;; SEQ ID NO 45
;; LENGTH: 35414
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-421-112-45

Query Match 96.2%; Score 845.4; DB 10; Length 35414;
Best Local Similarity 97.6%; Pred. No. 1.2e-274;
Matches 858; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TGTGGGAAAAGCAAGAGATCGATTGTTACTGTGTCTGTGTAGAAAAGAGATGACAT 60
DB 8681 TGTGGGAAAAGCAAGAGATCGATTGTTACTGTGTGTAGAAAAGAGATGACAT 8740
QY 61 AGGAGACTCATTTTGTATGTACTAAGAAAATTTCTTGCTTGAGATTCTGTATC 120
DB 8741 AGGAGACTCATTTTGTATGTACTAAGAAAATTTCTTGCTTGAGATTCTGTATC 8800
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAAATGTGTGTCTCACTCAGGGTTAA 180
DB 8801 TATGACCTTACCCCAACCCCGTCTCTGAAAATGTGTGTCTCACTCAGGGTTAA 8860
QY 181 ATGATTTAAGGGCGGTGAGATGTCTTTGTAAAGATGCTTGAAGAGACATGCTC 240
DB 8861 ATGATTTAAGGGCGGTGAGATGTCTTTGTAAAGATGCTTGAAGAGACATGCTC 8920
QY 241 CTTAAGAGTATCACTCCCTAATCTCAAGTACCAAGGAGACAAAACCTGGGAGG 300
DB 8921 CTTAAGAGTATCACTCCCTAATCTCAAGTACCAAGGAGACAAAACCTGGGAGG 8980
QY 301 CCGCAGGAGCTCTGCTAGGAAAAGCAGATTTGTCAAAGTTCTCCCATGTGATAG 360
DB 8981 CCGCAGGAGCTCTGCTAGGAAAAGCAGATTTGTCAAAGTTCTCCCATGTGATAG 9040
QY 361 CTTGAAATATGCGCTCGTGGGAAAGGAAAGATCGATCCCGACCCGACACCTGTA 420
DB 9041 CTTGAAATATGCGCTCGTGGGAAAGGAAAGATCGATCCCGACCCGACACCTGTA 9100
QY 421 AAGGCTGTGTGAGAGAGATTTAGTAAAGAGAGAAATGCTCTTGAGATTGAGACA 480
DB 9101 AAGGCTGTGTGAGAGAGATTTAGTAAAGAGAGAAATGCTCTTGAGATTGAGACA 9160
QY 481 AGAGAGAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 9161 AGAGAGAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9220
QY 541 GATTGTATGCTCATCTAGATAGAGAAAACCCGCTTGAAGGCTGAGAGTGAACCT 600
DB 9221 GATTGTATGCTCATCTAGATAGAGAAAACCCGCTTGAAGGCTGAGAGTGAACCT 9280
QY 601 GCGGGGAGCAATCTGCTTTGTAAAGCATGAGATGTTATGCTATGCAATCTAAAG 660
DB 9281 GCGGGGAGCAATCTGCTTTGTAAAGCATGAGATGTTATGCTATGCAATCTAAAG 9340
QY 661 CACAGCACTTAATCTTATGATGCTATGATGCAAGAACCTTGTGATGATGTTGCT 720
DB 9341 CACAGCACTTAATCTTATGATGCTATGATGCAAGAACCTTGTGATGATGTTGCT 9400
QY 721 GCTGACCTTCTCCCAATGCTTGTGACCTGACATCCCTCTTGTGAGAAACAC 780
DB 9401 GCTGACCTTCTCCCAATGCTTGTGACCTGACATCCCTCTTGTGAGAAACAC 9460
QY 781 CCAAGATGATCTAGTAAATCTAAGGAACTCAGAGGCTGGGGAGATCCCATATGCTG 840
DB 9461 CCAAGATGATCTAGTAAATCTAAGGAACTCAGAGGCTGGGGAGATCCCATATGCTG 9520
QY 841 AACGCTGTCCCGGGTCCCTTCTTCTCTATTA 879
|||||

DB 9521 AACGCTGTCCCGGGTCCCTTCTTCTCTATTA 9559

RESULT 4
US-10-067-514-1/c
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
; APPLICANT: Gensetec, Solvay
; APPLICANT: Janssen, Sif
; APPLICANT: Keynote, Sif
; TITLE OF INVENTION: HUMAN STROKE GENE
; FILE REFERENCE: 2345.2010-003
; CURRENT APPLICATION NUMBER: US/10/067,514
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 09/811/352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Human
US-10-067-514-1

Query Match 91.6%; Score 805.6; DB 14; Length 1691139;
Best Local Similarity 95.5%; Pred. No. 4.7e-260;
Matches 840; Conservative 0; Mismatches 39; Indels 1; Gaps 1;

QY 1 TGTGGGAAAAGCAAGAGATCGATTGTTACTGTGTCTGTGTAGAAAAGAGATGACAT 60
DB 116679 TGTGGGAAAAGCAAGAGATCGATTGTTACTGTGTGTAGAAAAGAGATGACAT 116620
QY 61 AGGAGACTCATTTTGTATGTACTAAGAAAATTTCTTGCTTGAGATTCTGTATC 120
DB 116619 AGGAGACTCATTTTGTATGTACTAAGAAAATTTCTTGCTTGAGATTCTGTATC 116560
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAAATGTGTGTCTCACTCAGGGTTAA 179
DB 116619 TATGACCTTACCCCAACCCCGTCTCTGAAAATGTGTGTCTCACTCAGGGTTAA 116500
QY 181 ATGATTTAAGGGCGGTGAGATGTCTTTGTAAAGATGCTTGAAGAGACATGCTC 239
DB 116649 ATGATTTAAGGGCGGTGAGATGTCTTTGTAAAGATGCTTGAAGAGACATGCTC 116440
QY 241 CTTAAGAGTATCACTCCCTAATCTCAAGTACCAAGGAGACAAAACCTGGGAGG 299
DB 116649 CTTAAGAGTATCACTCCCTAATCTCAAGTACCAAGGAGACAAAACCTGGGAGG 116380
QY 301 CCGCAGGAGCTCTGCTAGGAAAAGCAGATTTGTCAAAGTTCTCCCATGTGATAG 359
DB 116639 CCGCAGGAGCTCTGCTAGGAAAAGCAGATTTGTCAAAGTTCTCCCATGTGATAG 116320
QY 361 GCTGAAATATGCGCTCGTGGGAAAGGAAAGATCGATCCCGACCCGACACCCGT 419
DB 116619 GCTGAAATATGCGCTCGTGGGAAAGGAAAGATCGATCCCGACCCGACACCCGT 116260
QY 421 AAGGCTGTGTGAGAGATTTAGTAAAGAGAGAAATGCTCTTGAGATTGAGACA 479
DB 116629 AAGGCTGTGTGAGAGATTTAGTAAAGAGAGAAATGCTCTTGAGATTGAGACA 116200
QY 481 AAGAGAGAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 539
DB 116619 AAGAGAGAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 116140
QY 540 CGATTGTATCTCATATCTAGATAGAGAAAACCGCTTGAAGGCTGAGAGTGGAGAC 599
DB 116619 CGATTGTATCTCATATCTAGATAGAGAAAACCGCTTGAAGGCTGAGAGTGGAGAC 1166080
QY 600 TGTGGGAGCAATCTGCTTTGTAAAGCATGAGATTTATGTATGATCTTAA 659
DB 1166079 TGTGGGAGCAATCTGCTTTGTAAAGCATGAGATTTATGTATGATCTTAA 1166020
|||||

QY 660 GCACAGCACTTAATCCCTTACATTTGCTATGATGCAAAAGCCTTGTTCATGTTTGTG 719
 DB 1166019 GCACAGCACTTAATCCCTTACATTTGCTATGATGCAAAAGCCTTGTTCATGTTTGTG 1165960
 QY 720 TGTGACCCCTCTCCCAATTTGCTTGTATGACCCCTGACATATCCCTCTTTCAGAAACA 779
 DB 1165959 TGTGACCCCTCTCCCAATTTGCTTGTATGACCCCTGACATATCCCTCTTTCAGAAACA 1165900
 QY 780 CCACAGATGATGATTAATTAATTAAGGAAGTCAAGAGTGGGGAGTCCCTCATATGCT 839
 DB 1165899 CCACAGATGATTAATTAATTAAGGAAGTCAAGAGTGGGGAGTCCCTCATATGCT 1165840
 QY 840 GAACGCTGTTCCCGGGTCCCTCTTCTTCTCTATA 879
 DB 1165839 GAACGCTGTTCCCGGGTCCCTCTTCTTCTCTATA 1165800

RESULT 5
 US-10-419-723-1/c
 ; Sequence 1, Application US/10419723
 ; Publication No. US20040014099A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gretearsdotter, Solveig
 ; APPLICANT: Jonsdotter, Sig
 ; APPLICANT: Reymisdottir, Sigridur Th.
 ; APPLICANT: Thorleifsson, Gudmar
 ; APPLICANT: Gulcher, Jeffrey
 ; TITLE OF INVENTION: SUSCEPTIBILITY GENE FOR HUMAN STROKE;
 ; TITLE OF INVENTION: METHODS OF TREATMENT
 ; FILE REFERENCE: 2345.2010-005
 ; CURRENT APPLICATION NUMBER: US/10/419, 723
 ; PRIOR FILING DATE: 2003-04-18
 ; PRIOR APPLICATION NUMBER: 10/255,120
 ; PRIOR FILING DATE: 2002-09-25
 ; PRIOR APPLICATION NUMBER: 10/067,514
 ; PRIOR FILING DATE: 2002-02-04
 ; PRIOR APPLICATION NUMBER: 09/811,352
 ; PRIOR FILING DATE: 2001-03-19
 ; NUMBER OF SEQ ID NOS: 87
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 1691139
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-419-723-1

Query Match 91.6%; Score 805.6; DB 15; Length 1691139;
 Best Local Similarity 95.5%; Pred. No. 4,7e-260;
 Matches 840; Conservative 0; Mismatches 39; Indels 1; Gaps 1;
 QY 1 TGTGGGAAAAGCAAGAGATGATGTTTACTGTCTGTGTAGAAAAGATGACAT 60
 DB 1166679 TGTGGGAAAAGCAAGAGATGATGTTTACTGTCTGTGTAGAAAAGATGACAT 1166620
 QY 61 AGAGACTCCATTTTGTATGTAAGAAAATTTCTTGCCTTGAGATTCTGTATATC 120
 DB 1166619 AGAGACTCCATTTTGTATGTAAGAAAATTTCTTGCCTTGAGATTCTGTATATC 1166560
 QY 121 TATGACTTACCCCAACCCCGTCTCTGAAACAATGCTGTGTC-CACTCAGGGTTA 179
 DB 1166559 TATGACTTACCCCAACCCCGTCTCTGAAACAATGCTGTGTC-CACTCAGGGTTG 1166500
 QY 180 AATGATTAAGGGGCGTGCAGAGATGCTTGTAAACAGATGTTGAAGGACAGATGCT 239
 DB 1166499 AATGATTAAGGGGCGTGCAGAGATGCTTGTAAACAGATGTTGAAGGACAGATGCT 1166440
 QY 240 CCTTAAGATCATACCACTCCCTAATCTCAAGTACCAGGACACAAAACTGCGGAG 299
 DB 1166439 CCTTAAGATCATACCACTCCCTAATCTCAAGTACCAGGACACAAAACTGCGGAG 1166380
 QY 300 GCGGCAAGGACCTCTGCTGTGAAAGCAAGGATATTGTCACAGTTTCTCCCATGTGATA 359
 DB 1166379 GCGGCAAGGACCTCTGCTGTGAAAGCAAGGATATTGTCACAGTTTCTCCCATGTGATA 1166320

QY 360 GCGTGAATATGCGCTCGTGGAGAGGAAAAGACTTACCGTCCCGACCCGACCCGT 419
 DB 1166319 GCGTGAATATGCGCTCGTGGAGAGGAAAAGACTTACCGTCCCGACCCGACCCGT 1166260
 QY 420 AAGGGTCTGTGCTGAGAGAGATTACTAAAAGAGAAAGATTCCTTTCATTTGAGAC 479
 DB 1166259 AAGGGTCTGTGCTGAGAGAGATTACTAAAAGAGAAAGATTCCTTTCATTTGAGAC 1166200
 QY 480 AAGAGAGAGCATCTGTCTCTGCTGCGTCCCTGGGCAATGGAATGTCGCTATTAACC 539
 DB 1166199 AAGAGAGAGCATCTGTCTCTGCTGCGTCCCTGGGCAATGGAATGTCGCTATTAACC 1166140
 QY 540 CGATTGTATGCTTCATCTACTGATAGATAGGAAAACCGCTTAAAGGCTGGAGTGGACC 599
 DB 1166139 CGATTGTATGCTTCATCTACTGATAGATAGGAAAACCGCTTAAAGGCTGGAGTGGACA 1166080
 QY 600 TCGGGCAGCAATACCTGCTTTTGAAGCACTGAGATGTTATATGATATCTTAAA 659
 DB 1166079 TCGGGCAGCAATACCTGCTTTTGAAGCACTGAGATGTTATATGATATCTTAAA 1166020
 QY 660 GCACAGCACTTAATCCCTTACATTTGCTATGATGCAAAAGCCTTGTTCATGTTTGTG 719
 DB 1166019 GCACAGCACTTAATCCCTTACATTTGCTATGATGCAAAAGCCTTGTTCATGTTTGTG 1165960
 QY 720 TGTGACCCCTCTCCCAATTTGCTTGTATGACCCCTGACATCCCTCTTTCAGAAACA 779
 DB 1165959 TGTGACCCCTCTCCCAATTTGCTTGTATGACCCCTGACATCCCTCTTTCAGAAACA 1165900
 QY 780 CCACAGATGATGATTAATTAATTAAGGAAGTCAAGAGTGGGGAGTCCCTCATATGCT 839
 DB 1165899 CCACAGATGATTAATTAATTAAGGAAGTCAAGAGTGGGGAGTCCCTCATATGCT 1165840
 QY 840 GAACGCTGTTCCCGGGTCCCTCTTCTTCTCTATA 879
 DB 1165839 GAACGCTGTTCCCGGGTCCCTCTTCTTCTCTATA 1165800

RESULT 6
 US-10-233-958-15
 ; Sequence 15, Application US/10233958
 ; Publication No. US20040009468A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Mach, Bernard
 ; APPLICANT: Conrad, Bernard
 ; TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
 ; TITLE OF INVENTION: Thereof and Use in the Determination of Genetic
 ; TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
 ; FILE REFERENCE: 23135-504
 ; CURRENT APPLICATION NUMBER: US/10/233, 958
 ; PRIOR FILING DATE: 2002-09-03
 ; PRIOR APPLICATION NUMBER: 60/316,513
 ; PRIOR FILING DATE: 2001-08-31
 ; PRIOR APPLICATION NUMBER: 60/316,522
 ; PRIOR FILING DATE: 2001-08-31
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 15
 ; LENGTH: 975
 ; TYPE: DNA
 ; ORGANISM: Human endogenous retrovirus
 US-10-233-958-15
 Query Match 90.7%; Score 797.6; DB 15; Length 975;
 Best Local Similarity 94.9%; Pred. No. 2e-259;
 Matches 835; Conservative 0; Mismatches 44; Indels 1; Gaps 1;
 QY 1 TGTGGGAAAAGCAAGAGATGATGTTTACTGTCTGTGTAGAAAAGATGACAT 60
 DB 1 TGTGGGAAAAGCAAGAGATGATGTTTACTGTCTGTGTAGAAAAGATGACAT 60
 QY 61 AGAGACTCCATTTTGTATGTAAGAAAATTTCTTGCCTTGAGATTCTGTATATC 120

```

Db      61 AGGAGACTCCATTTTGTCTGTACTAGAAAATTTCTCTGCTGAGATCTGTATATC 120
Qy      121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTCTC-CACTCAGGGTTA 179
Db      121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTCTC-CACTCAGGGTTA 180
Qy      180 AATGATTTAAGGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGACAGATGCT 239
Db      181 AATGATTTAAGGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGACAGATGCT 240
Qy      240 CTTTAAAGTCAATCACTCTCTTAATCTCAAGTACCAGGAGACAAAACTCGGGAAG 299
Db      241 CATTAAAGTCAATCACTCTCTTAATCTCAAGTACCAGGAGACAAAACTCGGGAAG 300
Qy      300 GCGGCAAGGACCTCTGCTAGAAAAGCGAGTATTTGCCAAGTTTCTCCCAATGATA 359
Db      301 GCGGCAAGGACCTCTGCTAGAAAAGCGAGTATTTGCCAAGTTTCTCCCAATGATA 360
Qy      360 GCCTGAATATAGCGCTGTGGGAAAGGAAAGACCTGACCGTCCCGACCGGACCCGT 419
Db      361 GTCTGAATATAGCGCTGTGGGAAAGGAAAGACCTGACCGTCCCGACCGGACCCGT 420
Qy      420 AAAGGCTCTGTCTGAGAGGATTAGTAAAGAGAAAGAAATGCTCTTGGAGTTGAGAC 479
Db      421 AAAGGCTCTGTCTGAGAGGATTAGTAAAGAGAAAGAAATGCTCTTGGAGTTGAGAC 480
Qy      480 AAGAGAAAGGATCTGTCTCTGCTGCTGCTGCGGCAATGGAATGTCTCGGTATTAACC 539
Db      481 AAGAGAAAGGATCTGTCTCTGCTGCTGCTGCGGCAATGGAATGTCTCGGTATTAACC 540
Qy      540 CGATTGTATGCTCTCACTAGATAGGAAAGAAACCGCTTAGAGGCTGAGGTGGAGCC 599
Db      541 CGATTGTATGCTCTCACTAGATAGGAAAGAAACCGCTTAGAGGCTGAGGTGGAGCC 600
Qy      600 TGGCGGAGCAATATCTCTTTGTAAGACACTGAGATGTTATGTATGATCTATAA 659
Db      601 TGTGGGAGCAATATCTCTTTGTAAGACACTGAGATGTTATGTATGATCTATAA 660
Qy      660 GCACAGACATTAATCTTTTACATGTCTATGATGCAAAAGACCTTTGTCACATGTTGTC 719
Db      661 GCACAGACATTAATCTTTTACATGTCTATGATGCAAAAGACCTTTGTCACATGTTGTC 720
Qy      720 TGTGACCTCTCTCCCAATATGCTGTGACCGTGAACATCCCGCTTCGAGAAACA 779
Db      721 TGTGACCTCTCTCCCAATATGCTGTGACCGTGAACATCCCGCTTCGAGAAACA 780
Qy      780 CCCACAGATGATCAATTAATCTTAAGGAACTCAGAGGCTGGCGGATCTCCATATGCT 839
Db      781 CCCACAGATGATCAATTAATCTTAAGGAACTCAGAGGCTGGCGGATCTCCATATGCT 840
Qy      840 GAAAGCTGTGTTCCCGGGTCCCTTTCTTTCTCTATA 879
Db      841 GAAAGCTGTGTTCCCGGGTCCCTTTCTTTCTCTATA 880

```

```

RESULT 7
US-10-233-958-17
; Sequence 17, Application US/10233958
; Publication No. US20040009468A1
; GENERAL INFORMATION:
; APPLICANT: Mach, Bernard
; TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
; TITLE OF INVENTION: thereof and Use in the Determination of Genetic
; TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
; FILE REFERENCE: 23135-504
; CURRENT APPLICATION NUMBER: US/10/233,958
; PRIOR FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: 60/316,513
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/316,522

```

```

; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 17
; LENGTH: 975
; TYPE: DNA
; ORGANISM: Human endogenous retrovirus
US-10-233-958-17

Query Match      90.7%; Score 797.6; DB 15; Length 975;
Best Local Similarity 94.9%; Pred. No. 2e-259;
Matches 835; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

Qy      1 TGTGGGAAAAAGCAAGAGATCAGATTGTTACTGTGTCTGTGAGAAAGATAGACAT 60
Db      1 TGTGGGAAAAAGCAAGAGATGAGATGTTGTTACTGTGTCTGTATAGAAAGATAGACAT 60
Qy      61 AGGAGACTCCATTTTGTATGTACTAAGAAAATTTCTTGTGCTTGAATCTGTATATC 120
Db      61 AGGAGACTCCATTTTGTATGTACTAAGAAAATTTCTTGTGCTTGAATCTGTATATC 120
Qy      121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTCTC-CACTCAGGGTTA 179
Db      121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTCTC-CACTCAGGGTTA 180
Qy      180 AATGATTTAAGGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGACAGATGCT 239
Db      181 AATGATTTAAGGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGACAGATGCT 240
Qy      240 CTTTAAAGTCAATCACTCTCTTAATCTCAAGTACCAGGAGACAAAACTCGGGAAG 299
Db      241 CATTAAAGTCAATCACTCTCTTAATCTCAAGTACCAGGAGACAAAACTCGGGAAG 300
Qy      300 GCGGCAAGGACCTTGTCTGAGAAAGCGAGTATTTGCCAAGTTTCTCCCAATGATA 359
Db      301 ACCGCAAGGACCTTGTCTGAGAAAGCGAGTATTTGCCAAGTTTCTCCCAATGATA 360
Qy      360 GCGTGAATATAGCGCTGTGGGAAAGGAAAGACCTGACCGTCCCGACCGGACCCGT 419
Db      361 GTCTGAATATAGCGCTGTGGGAAAGGAAAGACCTGACCGTCCCGACCGGACCCGT 420
Qy      420 AAAGGCTCTGTCTGAGAGGATTAGTAAAGAGAAAGAAATGCTCTTGGAGTTGAGAC 479
Db      421 AAAGGCTCTGTCTGAGAGGATTAGTAAAGAGAAAGAAATGCTCTTGGAGTTGAGAC 480
Qy      480 AAGAGAAAGGATCTGTCTCTGCTGCTGCTGCGGCAATGGAATGTCTCGGTATTAACC 539
Db      481 AAGAGAAAGGATCTGTCTCTGCTGCTGCTGCGGCAATGGAATGTCTCGGTATTAACC 540
Qy      540 CGATTGTATGCTCTCACTAGATAGGAAAGAAACCGCTTAGAGGCTGAGGTGGAGCC 599
Db      541 CGATTGTATGCTCTCACTAGATAGGAAAGAAACCGCTTAGAGGCTGAGGTGGAGCC 600
Qy      600 TGGCGGAGCAATATCTCTTTGTAAGACACTGAGATGTTATGTATGATCTATAA 659
Db      601 TGTGGGAGCAATATCTCTTTGTAAGACACTGAGATGTTATGTATGATCTATAA 660
Qy      660 GCACAGACATTAATCTTTTACATGTCTATGATGCAAAAGACCTTTGTCACATGTTGTC 719
Db      661 GCACAGACATTAATCTTTTACATGTCTATGATGCAAAAGACCTTTGTCACATGTTGTC 720
Qy      720 TGTGACCTCTCTCCCAATATGCTGTGACCGTGAACATCCCGCTTCGAGAAACA 779
Db      721 TGTGACCTCTCTCCCAATATGCTGTGACCGTGAACATCCCGCTTCGAGAAACA 780
Qy      780 CCCACAGATGATCAATTAATCTTAAGGAACTCAGAGGCTGGCGGATCTCCATATGCT 839
Db      781 CCCACAGATGATCAATTAATCTTAAGGAACTCAGAGGCTGGCGGATCTCCATATGCT 840
Qy      840 GAAAGCTGTGTTCCCGGGTCCCTTTCTTTCTCTATA 879
Db      841 GAAAGCTGTGTTCCCGGGTCCCTTTCTTTCTCTATA 880

```

RESULT 8
US-10-233-958-2
; Sequence 2, Application US/10233958
; Publication No. US20040009468A1
; GENERAL INFORMATION:
; APPLICANT: Mach, Bernard
; APPLICANT: Conrad, Bernard
; TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
; TITLE OF INVENTION: thereof and Use in the Determination of Genetic
; TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
; FILE REFERENCE: 23135-504
; CURRENT APPLICATION NUMBER: US/10/233,958
; PRIOR FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: 60/316,513
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/316,522
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2689
; TYPE: DNA
; ORGANISM: Human endogenous retrovirus
US-10-233-958-2

Query Match 90.7%; Score 797.6; DB 15; Length 2689;
Best Local Similarity 94.9%; Pred. No. 3.8e-259;

Matches 835; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

```
QY 1 TGTGGGAAAAAGCAAGAGATGATGTTACTGTGTCTGTGTAGAAAGATGACAT 60
DB 1721 TGTGGGAAAAAGCAAGAGATGATGTTACTGTGTCTGTGTAGAAAGATGACAT 1780
QY 61 AGGAGATCTCATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 120
DB 1781 AGGAGATCTCATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1840
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTGTC-CACTCAGGGTTA 179
DB 1841 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTGTC-CACTCAGGGTTA 1900
QY 180 AATGATTTAAGGCGGTGACAGATGTCCTTGTAAACAGATGCTTGAAGGACATGCT 239
DB 1901 AATGATTTAAGGCGGTGACAGATGTCCTTGTAAACAGATGCTTGAAGGACATGCT 1960
QY 240 CCTTAAGATCATCACTCCCTAATCTCAAGTACCCAGGACACAAAACTCGGAG 299
DB 1961 CATTAAAGATCATCACTCCCTAATCTCAAGTACCCAGGACACAAAACTCGGAG 2020
QY 300 GCGGACGAGACCTCTGCTAGAAAGCAGGATTTGTCAACGTTTCCCATGATGATA 359
DB 2021 GCGGACGAGACCTCTGCTAGAAAGCAGGATTTGTCAACGTTTCCCATGATGATA 2080
QY 360 GCTGAATAATGCGCTGTGGGAAAGGAAAGACCTGACCGTCCCAAGCCGACCCGT 419
DB 2081 GCTGAATAATGCGCTGTGGGAAAGGAAAGACCTGACCGTCCCAAGCCGACCCGT 2140
QY 420 AAAGGCTCTGTCTGAGAGAGATTAATAAGAGAGAGAGAGAGAGAGAGAGAGAG 479
DB 2141 AAAGGCTCTGTCTGAGAGAGATTAATAAGAGAGAGAGAGAGAGAGAGAGAGAG 2200
QY 480 AAGAGAGAGAGATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 539
DB 2201 AAGAGAGAGAGATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2260
QY 540 CGATTGATCTCTCATCTACTGAGATAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 599
DB 2261 CGATTGATCTCTCATCTACTGAGATAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2320
QY 600 TGGGGGAGAGATCTGCTTTGTAAAGACTGAGATGTTATGTGTATGATATCTAAA 659
DB 600 TGGGGGAGAGATCTGCTTTGTAAAGACTGAGATGTTATGTGTATGATATCTAAA 659
```

```
DB 2321 TGTGGGAGCAATACCTGCTTTGTAAAGACTGAGATGTTATGTGTATGATATCTAAA 2380
QY 660 GCACAGACTTAATCTTTTACATTTGTATGATGATGATGATGATGATGATGATGATGAT 719
DB 2381 GCACAGACTTAATCTTTTACATTTGTATGATGATGATGATGATGATGATGATGATGAT 2440
QY 720 TGTGACCTCTCTCCCAACATTTGTTGACCTTGACATGCTGCTGCTGCTGCTGCTGCT 779
DB 2441 TGTGACCTCTCTCCCAACATTTGTTGACCTTGACATGCTGCTGCTGCTGCTGCTGCT 2500
QY 780 CCACAGATGATAGTAAATCTAAGGAGACTGAGAGGCTGGGGGATCTCCCATATGCT 839
DB 2501 CCACAGATGATAGTAAATCTAAGGAGACTGAGAGGCTGGGGGATCTCCCATATGCT 2560
QY 840 GAAAGCTGTTCCCGGGGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTATA 879
DB 2561 GAAAGCTGTTCCCGGGGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTATA 2600
```

RESULT 9
US-10-233-958-4
; Sequence 4, Application US/10233958
; Publication No. US20040009468A1
; GENERAL INFORMATION:
; APPLICANT: Mach, Bernard
; APPLICANT: Conrad, Bernard
; TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
; TITLE OF INVENTION: thereof and Use in the Determination of Genetic
; TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
; FILE REFERENCE: 23135-504
; CURRENT APPLICATION NUMBER: US/10/233,958
; PRIOR FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: 60/316,513
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/316,522
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2689
; TYPE: DNA
; ORGANISM: Human endogenous retrovirus
US-10-233-958-4

Query Match 90.7%; Score 797.6; DB 15; Length 2689;
Best Local Similarity 94.9%; Pred. No. 3.8e-259;

Matches 835; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

```
QY 1 TGTGGGAAAAAGCAAGAGATGATGTTACTGTGTCTGTGTAGAAAGATGACAT 60
DB 1721 TGTGGGAAAAAGCAAGAGATGATGTTACTGTGTCTGTGTAGAAAGATGACAT 1780
QY 61 AGGAGATCTCATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 120
DB 1781 AGGAGATCTCATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1840
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTGTC-CACTCAGGGTTA 179
DB 1841 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTGTC-CACTCAGGGTTA 1900
QY 180 AATGATTTAAGGCGGTGACAGATGTCCTTGTAAACAGATGCTTGAAGGACATGCT 239
DB 1901 AATGATTTAAGGCGGTGACAGATGTCCTTGTAAACAGATGCTTGAAGGACATGCT 1960
QY 240 CCTTAAGATCATCACTCCCTAATCTCAAGTACCCAGGACACAAAACTCGGAG 299
DB 1961 CATTAAAGATCATCACTCCCTAATCTCAAGTACCCAGGACACAAAACTCGGAG 2020
QY 300 GCGGACGAGACCTCTGCTAGAAAGCAGGATTTGTCAACGTTTCCCATGATGATA 359
DB 2021 ACCGACGAGACCTCTGCTAGAAAGCAGGATTTGTCAACGTTTCCCATGATGATA 2080
```



```

: NUMBER OF SEQ ID NOS: 46
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 26
: LENGTH: 6020
: TYPE: DNA
: ORGANISM: Human endogenous retrovirus
: US-10-233-958-26

```

Query Match	90.6%	Score 796	DB 15	Length 6020
Best Local Similarity	94.8%	Pred. NO.2.2e-258		
Matches 834	Conservative 0	Mismatches 45	Indels 1	Gaps 1

QY	1	IGTGGGGGAAAAGCAGAGAAGATCAGATTGTTACTGTGCTGTGTGTAAAGAAAGATGACAT	60
Db	2755	TGTGGGGGAAAAGCAGAGAAGATGAGATTGTTACTGTGCTGTATAGAAAGATGACAT	2696
QY	61	AGAGACTCCATTTTGTATGTACTAGAAAAAATCTTCTGCTTGAGATTCTGTTAAATC	120
Db	2695	AGGAGACTCCATTTTGTACTGTACTAGAAAAAATCTTCTGCTTGAGATTCTGTTAAATC	2633
QY	121	TATGACCTTAACCCCAACCCCGTGCTCTGTGAACAATGTGCTGTGTCTC-CATCTCAGGGTTA	179
Db	2635	TATGACCTTAACCCCAACCCCGTGCTCTGTGAACAATGTGCTGTGTCAATCACTCAGGGTTA	2576
QY	180	AATGGAATTAAAGGCGGCGTGACAGATGTGCTTTGTTAAACAGATGCTTGAAGCAGCATGCT	239
Db	2575	AATGGAATTAAAGGCGGCGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGCAGCATGCT	2511
QY	240	CCTTAAGAGTCATCAACACTCCCTTAATCTCAAGTACCCAGGGACACAAAATCTGGGAAG	299
Db	2515	CATTAAAGATCATCAACACTCCCTTAATCTCAAGTACCCAGGGACACAAAATCTGGGAAG	2456
QY	300	GCCGAGAGGACCTCTGCGCTAGGAAAGCCAGGATTTGTCCAAAGTTCTTCCCATGTGATA	359
Db	2455	ACCCGAGGACCTCTGCGCTAGGAAAGCTAAGATTTGTCCAAAGTTCTTCCCATGTGATA	2399
QY	360	GCTGAAATATGGCCTCGTGGGAAAGGAAAGACCTGACCTGTCCCCAGCCGACACCCGT	419
Db	2395	GTCGAAATATGGCCTCGTGGGAAAGGAAAGACCTGACCTGTCCCCAGACCAACCCGT	2336
QY	420	AAAGGCTGTGTGCTGAGAGAGATTGTTAAAGAGAAAGAAAGGCTCTTGCACTTGAGAC	479
Db	2335	AAAGGCTGTGTGCTGAGAGAGATTGTTAAAGAGAAAGAAAGGCTCTTGCACTTGAGAG	2276
QY	480	AAGAGAAAGGATCTGTCTCTGCTGCTGCTGCTGCGGCAATGGAATGCTCGATATAAAC	539
Db	2275	AAGAGAAAGATCTGTCTCTGCTGCTGCTGCTGCGGCAATGGAATGCTCGATATAAAC	2211
QY	540	CGATTGTATGCTCCATCTTACTGAGATAGGAAAAAACCGCTTAAAGGCTGAGAGTGGAC	599
Db	2215	CGATTGAAACATTCATCTTACTGAGATAGGAAAAAACCTGCTTAAAGGCTGAGAGTGGACA	2156
QY	600	TGCGGGGCGCAAAATCTGCTTTGTAAAGCACTGAGATGTTATGTATAGCATATCTAAAA	659
Db	2155	TGTGGGCGCAAAATCTGCTTTGTAAAGCACTTGAAGATGTTATGTATAGCATATCTAAAA	2099
QY	660	GCAACGACTTAATCTTTAATCTTGTCTATGTAGTGCAPAAAGACTTTGTTCAATGTTTGTCTC	719
Db	2095	GCACAGCACTTGATCTTTTACCTGTCTATGTAGTGCAPAAAGACTTTGTTTCACTGTTTGTCTC	2036
QY	720	TGCTGACCTCTCCCCACAAATTGTCTTGAGACCCGACACATCCCCCTCTTGAGAAAAA	779
Db	2035	TGCTCACCTCTCCCCACATATTGTCTTGAGACCCGACACATCCCCCTCTTGAGAAAAA	1976
QY	780	CCCAACAGATGATCACTAATAATACTAAGGAAACTCAAGAGCTGGCGGGATCTCTCATATGCT	839
Db	1975	CCCAACAGATGATCAATAATAATACTAAGGAAACTCAAGAGCTGGCGGGATCTCTCATATGCT	1911
QY	840	GAAGCGTGTCTCCCGGGTCCCTCTTCTTCTTCTCTATA 879	
Db	1915	GAAGCGTGTGTCTCCCGGGCCCTTATATCTTCTTCTCTATA 1876	

```

RESULT 12
US-10-233-958-42
; Sequence 42, Application US/10233958
; Publication No. US20040009468A1
; GENERAL INFORMATION:
; APPLICANT: Mach, Bernard
; TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
; TITLE OF INVENTION: Theoret and Use in the Determination of Genetic
; TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
; FILE REFERENCE: 23135-504
; CURRENT APPLICATION NUMBER: US/10/233,958
; CURRENT FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: 60/316,513
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/316,522
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42
; LENGTH: 10569
; TYPE: DNA
; ORGANISM: Human endogenous retrovirus
US-10-233-958-42

```

Query Match	90.6%	Score 796	DB 15	Length 10569
Best Local Similarity	94.8%	Pred. No. 3.2e-258		
Matches 834	0	Mismatches 45	Indels 1	Gaps 1

QY	1	TGTGGGAAAAGCAAGAGATCAGATTGTACTGTGTCTGTGTGAAGAAGATGACAT	60
Db	765	TGTGGGAAAAGCAAGAGAGTCAGATTGTTACTGTGTCTGTATATGAAGAAGTACAT	824
QY	61	AGGAGACTCCATTTTGTATTATCTACGAAAAAATTTCTTGCCTTGAGATTCTGTTAATC	120
Db	825	AGGAGACTCCATTTTGTATTCTACTAGAAAAAATTAATTTCTGCTTGAGATGCTGTTAATC	884
QY	121	TATGACCTTACCCCCAACCCCGTCTCTGTAAACAATGTGCTGTGTC-CACTCAGGGTTA	179
Db	885	TATGACCTTACCCCCAACCCCGTCTCTGTAAACAATGTGCTGTGTCACAACTCAGGGTTA	944
QY	180	AATGATTAAAGGCGGTGACAGATGTGTGTTTAAACAGATGCTTGAAGCAGCATCT	239
Db	945	AATGATTAAAGGCGGTGACAGATGTGTGTTTAAACAGATGCTTGAAGCAGCATCT	1004
QY	240	CCTTAGAGTCATCACTCTCTTAATCTCAAGTACCAGGACACAAAACTGCGGANG	299
Db	1005	CATTAGAGTCATCACTCTCTTAATCTCAAGTACCAGGACACAAAACTGCGGANG	1064
QY	300	GCCGAGGGACCTCTGCCTTAGAAAAAGCAGTATTGTCCAAAGTTTCTCCCATGTGATA	359
Db	1065	CCTGAGGGGCTCTGCCTTAGAAAAAGCAGTATTGTCCAAAGTTTCTCCCATGTGATA	1122
QY	360	GCTGAATATATGAGCTCGTGGGAAAGGAAAGACCTGACCGTCCCCACCCGACACCGGT	419
Db	1125	GTCGAAATATATGAGCTCGTGGGAAAGGAAAGACCTGACCGTCCCCACCCGACACCGGT	1188
QY	420	AAAGGCTGTGTCTGAGAGGATTATGTAAGAAGAGAAATGCTCTTTCGACTTGAAC	479
Db	1185	AAAGGCTGTGTCTGAGAGGATTATGTAAGAAGAGAAATGCTCTTTCGACTTGAAC	1244
QY	480	AAAGGAAGGATCTGCTCTGCGTGTCCCTGGGCAATGGAATGTCGGATATAAAC	539
Db	1245	AAAGGAAGGATCTGTTTCCCGCCATCTCTGGGCAATGGAATGTCGGATATAAAC	1304
QY	540	CGATTGTATGCTCATCTACTAGATAGGAAAAACCGCTTAGGGCTGTGAGTGGAGCC	599
Db	1305	CGATTGTATGCTTCCACTCTAGATAGGAAAAACCACTTTAAGGCTGTGAGTGGAGCA	1366
QY	600	TGCGGGCAGCAATCTGCTTTGTAAAGCACTGAGATGTTATGTGTATGCAATCTTAAA	659
Db	1365	TGCGGGCAGCAATCTGCTTTGTAAAGCACTGAGATGTTATGTGTATGCAATCTTAAA	1422

QY 660 GCACAGACTTAATCCTTACATGTCTATGATGCAAGACCTTGTTCACATGTTGTC 719
DB 1425 GCACAGACTTAATCCTTACATGTCTATGATGCAAGACCTTGTTCACATGTTGTC 1484
QY 720 TGCTGACCTCTCTCCCAACATTTGTTGTGACCTGACACATCCCTCTTGGAAACA 779
DB 1485 TGCTGACCTCTCTCCCAACATTTGTTGTGACCTGACACATCCCTCTTGGAAACA 1544
QY 780 CCACAGATATATCTAATTAATCTAAGGAACTCAGAGCTGCGGATCCCTCATATGCT 839
DB 1545 CCACAGATATATCTAATTAATCTAAGGAACTCAGAGCTGCGGATCCCTCATATGCT 1604
QY 840 GAACGCTGTTCCTCCGCGCTCTTCTTCTCTATA 879
DB 1605 GAACGCTGTTCCTCCGCGCTCTTCTTCTCTATA 1644

RESULT 13
US-10-233-958-18
; Sequence 18, Application US/10233958
; Publication No. US2004009468A1
; GENERAL INFORMATION:
; APPLICANT: Mach, Bernard
; TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
; TITLE OF INVENTION: Thereof and Use in the Determination of Genetic
; TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
; TITLE OF INVENTION: Provirus
; FILE REFERENCE: 23135-504
; CURRENT APPLICATION NUMBER: US/10/233,958
; PRIOR FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: 60/316,513
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/316,522
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 18
; LENGTH: 969
; TYPE: DNA
; ORGANISM: Human endogenous retrovirus
US-10-233-958-18

Query Match 90.4%; Score 794.4; DB 15; Length 969;
Best Local Similarity 94.7%; Pred. No. 2.4e-258;
Matches 833; Conservative 0; Mismatches 46; Indels 1; Gaps 1;

QY 1 TGTTGGGAAAAGCAAGAGATCAGATTGTACTGTGTCTGTAGAAAGATAGACAT 60
DB 1 TGTTGGGAAAAGCAAGAGATCAGATTGTACTGTGTCTGTAGAAAGATAGACAT 60
QY 61 AGGAGACTCATTGTTGTTATGTAAGAAAAATTTCTGCTTGAGATTCTGTTATC 120
DB 61 AGGAGACTCATTGTTGTTATGTAAGAAAAATTTCTGCTTGAGATTCTGTTATC 120
QY 121 TATGACTTATCCCAACCCCGTCTCTGAAAACATGTGCTGTGTC-CACTCAGGGTTA 179
DB 121 TATGACTTATCCCAACCCCGTCTCTGAAAACATGTGCTGTGTC-CACTCAGGGTTA 180
QY 180 AATGATTAAAGGGGGGAGAGATGTGTTGTTAAACAGATGCTTGAAGGACACATGCT 239
DB 181 AATGATTAAAGGGGGGAGAGATGTGTTGTTAAACAGATGCTTGAAGGACACATGCT 240
QY 240 CTTTAAGATCATCACCACCTCCCTAATCTCAATGACAGGAGACAAAACTCGGAG 299
DB 241 CATTAAAGATCATCACCACCTCCCTAATCTCAATGACAGGAGACAAAACTCGGAG 300
QY 300 GCGCAGGAGACTCTGCTCTAGAAAAGCAGATATTTGTCACAGTTTCTCCCATGTGATA 359
DB 301 CTGTCAGGGGGCTCTGCTCTAGAAAAGCAGATATTTGTCACAGTTTCTCCCATGTGATA 360
QY 360 GCGTGAATATGCGCTGTGGGAAAGGAAAGACCTGACCGTCCCGACCGCACCGCT 419

DB 361 GTCTGAATATGCGCTGTGGGAAAGGAAAGACCTGACCGTCCCGACCGACACCGCT 420
QY 420 AAGGGCTGTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTTGACATGAGAC 479
DB 421 AAGGGCTGTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTTGACATGAGAC 480
QY 480 AAGAGAGAGCATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 539
DB 481 AAGAGAGAGCATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 540 CGATTGATGCTCCATCTACTAGATAGGAAAAACCGCTTGAAGGCTGAGGTGGAC 599
DB 541 CGATTGATGCTCCATCTACTAGATAGGAAAAACCGCTTGAAGGCTGAGGTGGAC 600
QY 600 TGCGGAGAGAAATCTCTCTTGTAAACACTGAGATGTTATATGATATATCTAAA 659
DB 601 TGCGGAGAGAAATCTCTCTTGTAAACACTGAGATGTTATATGATATATCTAAA 660
QY 660 GCACAGACTTAATCCTTACATGTCTATGATGCAAGACCTTGTTCACATGTTGTC 719
DB 661 GCACAGACTTAATCCTTACATGTCTATGATGCAAGACCTTGTTCACATGTTGTC 720
QY 720 TGCTGACCTCTCTCCCAACATTTGTTGTGACCTGACACATCCCTCTTGGAAACA 779
DB 721 TGCTGACCTCTCTCCCAACATTTGTTGTGACCTGACACATCCCTCTTGGAAACA 780
QY 780 CCACAGATATATCTAATTAATCTAAGGAACTCAGAGCTGCGGATCCCTCATATGCT 839
DB 781 CCACAGATATATCTAATTAATCTAAGGAACTCAGAGCTGCGGATCCCTCATATGCT 840
QY 840 GAACGCTGTTCCTCCGCGCTCTTCTTCTCTATA 879
DB 841 GAACGCTGTTCCTCCGCGCTCTTCTTCTCTATA 880

RESULT 14
US-10-233-958-16
; Sequence 16, Application US/10233958
; Publication No. US2004009468A1
; GENERAL INFORMATION:
; APPLICANT: Mach, Bernard
; TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
; TITLE OF INVENTION: Thereof and Use in the Determination of Genetic
; TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
; TITLE OF INVENTION: Provirus
; FILE REFERENCE: 23135-504
; CURRENT APPLICATION NUMBER: US/10/233,958
; PRIOR FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: 60/316,513
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/316,522
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 16
; LENGTH: 975
; TYPE: DNA
; ORGANISM: Human endogenous retrovirus
US-10-233-958-16

Query Match 90.4%; Score 794.4; DB 15; Length 975;
Best Local Similarity 94.7%; Pred. No. 2.4e-258;
Matches 833; Conservative 0; Mismatches 46; Indels 1; Gaps 1;

QY 1 TGTTGGGAAAAGCAAGAGATCAGATTGTACTGTGTCTGTAGAAAGATAGACAT 60
DB 1 TGTTGGGAAAAGCAAGAGATCAGATTGTACTGTGTCTGTAGAAAGATAGACAT 60
QY 61 AGGAGACTCATTGTTGTTATGTAAGAAAAATTTCTGCTTGAGATTCTGTTATC 120
DB 61 AGGAGACTCATTGTTGTTATGTAAGAAAAATTTCTGCTTGAGATTCTGTTATC 120

QY	121	TATGACCTTAACCCCAACCCCGTCTCTGTAAACATGTGCTGTGTC-CACTCAGGGTTA	179
Db	121	TATGACCTTAACCCCAACCCCGTCTCTGTAAACATGTGCTGTGTC-CACTCAGGGTTA	180
QY	180	AATGATTAAGGGCGGTGCAAGATGTCTTTGTTAAACAATGCTTGAAGCAATGCT	229
Db	181	AATGATTTAAGGGGTGTGCAAGATGTGCTTTGTTTAAACAGATGCTTGAAGCAATGCT	240
QY	240	CCCTAAGAGTCATCAACACTCCCTTAATCTCAAGTACCCAGGGACACAAAATGCGGAAG	239
Db	241	CATTAAAGAGTCATCAACACTCCCTTAATCTCAAGTACCCAGGGACACAAAATGCGGAAG	300
QY	300	GCCGCAAGGACCTCTGCTCCTTAGGAAAGCCAGATATTGTCCAAAGTTTCTCCCATGTGATA	359
Db	301	GCCGCAAGGACCTCTGCTCCTTAGGAAAGCCAGATATTGTCCAAAGTTTCTCCCATGTGATA	360
QY	360	GCCCTGAATATGSGCTCTGTGGAGAGGAAAGACTGACCGTCTCCCAAGCCGACACCCGT	419
Db	361	GTCTGAAATATGSGCTCTGTGGAGAGGAAAGACTGACCAATCCCAAGACCAACCCGT	420
QY	420	AAAGGGTCTGTGTGAGAGAAATTGTTAAABAAGAAAGAAATGCTCTGTGATGAGAC	479
Db	421	AAAGGGTCTGTGTGAGAGAAATTGTTAAABAAGAAAGAAATGCTCTGTGATGAGAG	480
QY	480	AAAGAGAAAGCATCTTCTCTGCTGTCTCGGGCAATGGAATGTCTCGTATATAAAC	539
Db	481	AAAGAGAAAGCATCTGTCTCTGCTGTCTCGGGCAATGGAATGTCTCAATATAAAC	540
QY	540	CGATTGTATGCTTCATCTCTGTAGATAGGAAAAACCGCTTAAGGGCTGAGGTGGAC	599
Db	541	CGATTGAACATTCCTCACTCACTAGATAGGAAAAACCTCGCTTAAGGGCTGAGGTGGACA	600
QY	600	TGCGGGGAGCAATACCTGCTTTGTTAAAGCACTGAGATGTTTATGTATGATCATCTATAA	659
Db	601	TGTGGGACGAATACCTGCTTTGTTAAAGCACTGAGATGTTTATGTATGATCATCTATAA	660
QY	660	GCAACAGCATTAATCCTTTACATTTGTCTATGATGCAAGACCTTTGTTCAATGTTTGTG	719
Db	661	GCAACAGCATTAATCCTTTACATTTGTCTATGATGCAAGACCTTTGTTCAATGTTTGTG	720
QY	720	TGCTGACCTCTCTCCCAACAATTTGTCTTGTGACCCCTGACACATCCCTCTTCGAGAAACA	779
Db	721	TGCTGACCTCTCTCCCAACAATTTGTCTTGTGACCCCTGACACATCCCTCTTCGAGAAACA	780
QY	780	CCACACGATGATCAGTAAATACTAAAGGGAACCTCAGAGGCTGGCGGGAATCCCTCATATGCT	839
Db	781	CCACACGATGATCAGTAAATACTAAAGGGAACCTCAGAGGCTGGCGGGAATCCCTCATATGCT	840
QY	840	GAAAGCTGTGTTCCCGGGGTCCCTTCTTTCTTTCTCTATA	879
Db	841	GAAAGCTGTGTTCCCGGGGTCCCTTCTTTCTTTCTCTATA	880
RESULT 15			
US-10-233-958-19			
: Sequence 19, Application US/10233958			
: Publication No. US20040009468A1			
: GENERAL INFORMATION:			
: APPLICANT: Mach, Bernard			
: TITLE OF INVENTION: Allelic Variants of HER V-Ki8, Method for the Analysis			
: TITLE OF INVENTION: Theoreof and Use in the Determination of Genetic			
: TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-Ki8			
: FILE REFERENCE: 23135-504			
: CURRENT APPLICATION NUMBER: US/10/233,958			
: PRIOR FILING DATE: 2002-09-03			
: PRIOR APPLICATION NUMBER: 60/316,513			
: PRIOR FILING DATE: 2001-08-31			
: PRIOR APPLICATION NUMBER: 60/316,522			
: PRIOR FILING DATE: 2001-08-31			
: NUMBER OF SEQ ID NOS: 46			

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
;
; LENGTH: 1010
;
; TYPE: DNA
;
; ORGANISM: Human endogenous retrovirus
;
US-10-233-958-19

```

Query Match	90.2%;	Score 792.8;	DB 15;	Length 1010;
Best Local Similarity	94.5%;	Pred. No. 8.6e-258;		
Matches 832; Conservative	0;	Mismatches 47;	Indels 1;	Gaps 1;

QY	1	1G1GAGGAGAGCAGAGAGAG1CAGAT1G1TAC1G1G1G1G1AGAGAGAG1AGACAT	80
DY	1	TGTGGGGAGAGCAACAGAGTCAGATTGTATTCGTGTCTGTATAGAAAGAAATAGACAT	60
QY	61	AGGAGACTCCATTTTGTATATGACTAAGAAAAATTTCTTGCCCTTGAGATCTGTATATC	120
DY	61	AGGAGACTCCATTTTGTCTGTACTAAGAAAAATTTATTCCTTGAGATCTGTATATC	120
QY	121	TATGACCTTACCCCCCAACCCCGTGTCTTGAAACATGTGTGTGTC-CACTCAGGGTTA	179
DY	121	TATGACCTTACCCCCCAACCCCGTGTCTTGAAACATGTGTGTGTGTC-TCAACAGGGTTA	180
QY	180	AATGGAATTAAGGGGGGTGCAAGATGTGTCTTTGTAAACAGATGCTTGAGGACAGATGCT	239
DY	181	AATGGAATTAAGGGGGGTGCAAGATGTGTCTTTGTAAACAGATGCTTGAGGACAGATGCT	240
QY	240	CCTTAAAGATCATCACACTCCCTAATCTCAAGTACCAGGACACAAAACTGCGAAG	299
DY	241	CATTAAAGTATATCACACTCCCTAATCTCAAGTACCAGGACACAAAACTGCGAAG	300
QY	300	GCCGACGGACCTTGCTTGAGAAAACGACGATATTTGCCAGTTTCTCCCATGTGATA	359
DY	301	GCTGCAAGGGGCTTGCTTGAGAAAACGAGTATATTTGCCAGTTTCTCCCATGTGAGA	360
QY	360	GCCGAAATATAGGCTGTGTGGGAAAGGAAAGACCTGACCGTCCCGACCGGACACCGGT	419
DY	361	GTCGAAATATAGGCTGTGTGGGAAAGGAAAGACCTGACCGTCCCGACCGGACACCGAT	420
QY	420	AAAAGGATCTGTGTGAGAGGATTAATAGTAAAGAGAAAGATTGCTCTTGCAATTTGAGAC	479
DY	421	AAAAGGATCTGTGTGAGAGGATTAATAGTAAAGAGAAAGATTGCTCTTGCAATTTGAGAC	480
QY	480	AAGAGAAAGGATTTGTCTCTGCTGCTGCTGCGGCAATGGAATGTCTCGTATATAAAC	539
DY	481	AAGAGAAAGGATTTGTCTCTGCTGCTGCTGCGGCAATGGAATGTCTCGTATATAAAC	540
QY	540	CGATTGATGCTCCATCTACTGAGATAGAGGAAAAACCGCTTAGGGCTGAGAGTGGAGCC	599
DY	541	CGATTGATGCTCCATCTACTGAGATAGAGGAAAAACCGCTTAGGGCTGAGAGTGGAGCA	600
QY	600	TGCGGGCAGCAATATCTGCTTTGTAAAGCACTGAGATGTTATGTGTATGATATCTAAAA	659
DY	601	TGCGGGCAGCAATATCTGCTTTGTAAAGCACTGAGATGTTATGTGTATGATATCTAAAA	660
QY	660	GCACAGACTTAATCCTTTACATGTGTATAGATGCAAAAGACCTTTGTTCACATGTTTGT	719
DY	661	GCACAGACTTAATCCTTTACATGTGTATAGATGCAAAAGACCTTTGTTCACATGTTTGT	720
QY	720	TGCTGACCTCTCCCAACAATGTCTTGAGACCTGACACATCCCGCTCTCGAAGAAACA	779
DY	721	TGCTGACCTCTCCCAACAATGTCTTGAGACCTGACACATCTCCCTCTCAGAGAAACA	780
QY	780	CCACAGATGATCGTAAATATCTAAGGAACTCAGAGGCTGGCGGGATCTTCATATGCT	839
DY	781	CCACAGATGATCGTAAATATCTAAGGAGGACTCAGAGGCTGGGGATCTTCATATGCT	840
QY	840	GAAAGCTGTGTTCCCGGGTCCCTTTCTTTCTTTCTCATATA	879
DY	841	GAAAGCTGTGTTCCCGGGTCCCTTTCTTTCTTTCTCATATA	880

✓ Mon Mar 1 09:41:06 2004

Job time : 393.694 secs

us-10-016-604-5.rnpb

Page 11

This Page Blank (us010)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2004, 21:22:12 ; Search time 457.012 Seconds
(without alignments)
8170.833 Million cell updates/sec

Title: US-10-016-604-5

Perfect score: 879

Sequence: 1 tctgggggaaagacagagagag.....cccttccttccttcctcata 879

Scoring table: IDENTITY_NUC
Gap 10.0, Gape 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: N_Geneseq_29Jan04:*
2: geneseqn1980s:*
3: geneseqn1980s:*
4: geneseqn2000s:*
5: geneseqn2001as:*
6: geneseqn2002as:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004as:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	879	100.0	879	ABX04572	Abx04572 Human end
2	879	100.0	879	ADCI16258	Adci16258 Human her
3	879	100.0	968	ABX04691	Abx04691 Human end
4	879	100.0	968	ADCI15768	Adci15768 LTR of he
5	856.6	97.5	32249	ABX05336	Abx05336 Human rep
6	856.6	97.5	32249	ABX05336	Abx05336 Human tes
7	855	97.3	9179	ABX04612	Abx04612 Human end
8	851.8	96.9	968	ABX04693	Abx04693 Human end
9	851.8	96.9	968	ADCI15770	Adci15770 HML-2 LTR
10	845.4	96.2	35414	ABX04147	Abx04147 TR12 rela
11	840.6	95.6	222930	ABX04349	Abx04349 Human end
12	837.4	95.3	968	ABX04695	Abx04695 Human end
13	837.4	95.3	968	ABX04694	Abx04694 Human end
14	837.4	95.3	968	ADCI15772	Adci15772 HML-2 LTR
15	837.4	95.3	968	ADCI15771	Adci15771 HML-2 LTR
16	835.4	95.0	6556	AAK70301	Aak70301 Human imm
17	834.4	94.9	96599	AAK70301	Aak70301 Human pro
18	834.4	94.9	96599	ADCI15773	Adci15773 Human PRL
19	834.4	94.9	96599	ADCI15774	Adci15774 Human PRL
20	832.2	94.7	2431	AAK94873	Aak94873 Human ful
21	824.2	93.8	154902	ABX08198	Abx08198 Human ost
22	813.6	92.6	139904	ABX083562	Abx083562 Human CDN
23	805.6	91.6	110000	ABX08336_11	Continuation (12 o

24	804	91.5	8095	4	AAK82357	Aak82357 Human imm
25	804	91.5	11122	6	ABX04611	Abx04611 Human end
26	797.6	90.7	975	7	ACC43225	Acc43225 Nucleotid
27	797.6	90.7	975	7	ACC43227	Acc43227 Nucleotid
28	797.6	90.7	2689	7	ACC43221	Acc43221 Nucleotid
29	797.6	90.7	2689	7	ACC43223	Acc43223 Nucleotid
30	797.6	90.7	1701	5	AAK83303	Aak83303 DNA encod
31	797	90.7	1826	5	AAK83303	Aak83303 DNA encod
32	796	90.6	972	7	ACC43230	Acc43230 Nucleotid
33	796	90.6	6020	7	ACC43236	Acc43236 Nucleotid
34	794.4	90.4	969	7	ACC43228	Acc43228 Nucleotid
35	794.4	90.4	975	7	ACC43226	Acc43226 Nucleotid
36	792.8	90.2	1010	7	ACC43229	Acc43229 Nucleotid
37	789.2	89.8	144792	9	ADCI15769	Adci15769 Human GPC
38	789.2	89.8	349981	9	ADCI15769	Adci15769 Human GPC
39	787.8	89.6	962	6	ABX04692	Abx04692 Human end
40	787.8	89.6	962	6	ADCI15769	Adci15769 HML-2 LTR
41	784	89.2	198285	6	ABX04699	Abx04699 Human CDN
42	784	89.2	198285	6	ABN97319	Abn97319 Gene #381
43	778.6	88.6	1574	4	AAI99316	Aai99316 Human exc
44	778.6	88.6	1574	5	AAI63666	Aai63666 Human kid
45	766.2	87.2	1782	5	AAK79381	Aak79381 DNA encod

ALIGNMENTS

RESULT 1	ABX04572 standard; DNA; 879 BP.
ID	ABX04572
AC	ABX04572;
DT	14-JAN-2003 (first entry)
XX	
XX	Human endogenous retrovirus k (herv-k) U3R region.
XX	
XX	Human endogenous retrovirus; herv; prostate cancer; testicular cancer; multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease; cancer; transgenic animal; ds.
XX	
XX	Human endogenous retrovirus.
XX	
XX	MO200246477-A2.
XX	
XX	13-JUN-2002.
XX	
XX	07-DEC-2001; 2001WO-US047824.
XX	
XX	07-DEC-2000; 2000US-0251830P.
XX	
XX	07-DEC-2001; 2001US-00016604.
XX	
XX	(CHIR) CHIRON CORP.
XX	
XX	Garcia P, Hardy SF, Williams LT, Escobedo J;
XX	WPI; 2002-691475/74.
XX	
XX	Novel isolated polypeptides useful for diagnosis of prostate cancer.
XX	
XX	Claim 4; Page 129; 152pp; English.
XX	
XX	The invention describes novel isolated polypeptides (I, Ib) useful for diagnosing prostate cancer comprising obtaining a patient sample containing prostate cells and detecting the presence or absence of an expression product of a HML-2 endogenous retrovirus in a patient sample. Polymorphisms associated with (I) are useful for diagnosis or treatment of testicular cancer, multiple sclerosis or insulin-dependent diabetes mellitus. An inhibitor of a HML-2 protease and a transdominant negative mutant of HML-2 CORP are also useful in the manufacture of a medicament for treating prostate cancer. (I) and (Ib) are useful for generating antibodies specific to the polypeptides associated with cancer, as targets for therapeutic intervention, and in immunising a transgenic

CC animal. This sequence represents a region of human endogenous retrovirus
CC (herv) of the HML-2 sub-group
XX
SQ Sequence 879 BP; 231 A; 212 C; 211 G; 225 T; 0 U; 0 Other;

Query Match 100.0%; Score 879; DB 6; Length 879;
Best Local Similarity 100.0%; Pred. No. 4.3e-288;
Matches 879; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TGTGGGAAAAAGCAAGAGATGATGTTACTGTCGTCTGTAGAAAGAGATGACAT 60
DB 1 TGTGGGAAAAAGCAAGAGATGATGTTACTGTCGTCTGTAGAAAGAGATGACAT 60
QY AGAGACTCCATTTTGTATGATCTAAGAAAAATTTCTTCGCTTGAAGATTTGTAATC 120
DB AGAGACTCCATTTTGTATGATCTAAGAAAAATTTCTTCGCTTGAAGATTTGTAATC 120
QY 121 TATGACCTTACCCCAACCCCGTCTCTGTAAACATGTGCTGTGCACTCAGGGTTAA 180
DB 121 TATGACCTTACCCCAACCCCGTCTCTGTAAACATGTGCTGTGCACTCAGGGTTAA 180
QY 181 ATGGATTAAAGGGGCGTGAGAGATGTCCTTGTAAACAGATGCTTGAAGGAGATGCTC 240
DB 181 ATGGATTAAAGGGGCGTGAGAGATGTCCTTGTAAACAGATGCTTGAAGGAGATGCTC 240
QY 241 CTTAAGAGTCATCAACATCCCTATCTCAAGTACCAGGAGACAAAACTGCGGAAG 300
DB 241 CTTAAGAGTCATCAACATCCCTATCTCAAGTACCAGGAGACAAAACTGCGGAAG 300
QY 301 CCGGAGGAGCTCTGCTTGAAGAAACCAAGTATGTTCCACGTTTCTCCATGTGATAG 360
DB 301 CCGGAGGAGCTCTGCTTGAAGAAACCAAGTATGTTCCACGTTTCTCCATGTGATAG 360
QY 361 CCGGAGGAGCTCTGCTTGAAGAAACCAAGTATGTTCCACGTTTCTCCATGTGATAG 360
DB 361 CCGGAGGAGCTCTGCTTGAAGAAACCAAGTATGTTCCACGTTTCTCCATGTGATAG 360
QY 421 AAGGCTCTGTCGTCAGAGAGATTTAGTAAAGAGAGAAATGCTCTTGCAGTTGAGACA 480
DB 421 AAGGCTCTGTCGTCAGAGAGATTTAGTAAAGAGAGAAATGCTCTTGCAGTTGAGACA 480
QY 481 AGAGAAAGGAGTCGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 481 AGAGAAAGGAGTCGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 541 GATTGTATGCTCCATCTAAGATAGAGAAACCCGCTTAAAGGCTGAGAGTGGACCT 600
DB 541 GATTGTATGCTCCATCTAAGATAGAGAAACCCGCTTAAAGGCTGAGAGTGGACCT 600
QY 601 GCGGGCAGCAATATCTGCTTGTAAAGCACTGAGATGTTATGTATGATATCTAAAG 660
DB 601 GCGGGCAGCAATATCTGCTTGTAAAGCACTGAGATGTTATGTATGATATCTAAAG 660
QY 661 CACAGCACTTAATCTTTTACATGTCTATGATGCAAAACCTTGTTCACATGTTTGTCT 720
DB 661 CACAGCACTTAATCTTTTACATGTCTATGATGCAAAACCTTGTTCACATGTTTGTCT 720
QY 721 GCTGACCTCTCCCAACATTTGTGTGACCTTGAACATCCCTCTTCTGAGAAACAC 780
DB 721 GCTGACCTCTCCCAACATTTGTGTGACCTTGAACATCCCTCTTCTGAGAAACAC 780
QY 781 CCACAGATGATCACTAATAATCTAAGGAACTCAGAGGCTGCGGGATCTCATATGCTG 840
DB 781 CCACAGATGATCACTAATAATCTAAGGAACTCAGAGGCTGCGGGATCTCATATGCTG 840
QY 841 AAGCTGTGTCCTCCGCGGTCCCTTCTTCTCTCTATA 879
DB 841 AAGCTGTGTCCTCCGCGGTCCCTTCTTCTCTCTATA 879
```

RESULT 2
ADCl6258
ID ADCl6258 standard; DNA; 879 BP.

```
XX AC ADCl6258;
XX 18-DEC-2003 (first entry)
XX DE Human herv-k(hml)-2(hom) U3R region nucleotide sequence SEQ ID NO:5.
XX KW prostate cancer; HML-2; env; cytosolic; neuroprotective; antidiabetic;
XX KM immunostimulant; vaccine; gene therapy; breast cancer; testicular cancer;
XX multiple sclerosis; insulin-dependent diabetes mellitus; human; gene; ds.
XX OS Homo sapiens.
XX PN WO2003050258-A2.
XX PD 19-JUN-2003.
XX PE 09-DEC-2002; 2002MO-US039344.
XX PR 07-DEC-2001; 2001US-00016604.
XX PR 07-DEC-2001; 2001US-0340064P.
XX PR 07-DEC-2001; 2001WO-US047824.
XX PR 12-JUN-2002; 2002US-0388046P.
XX PA (CHIR ) CHIRON CORP.
XX PI Garcia P, Hardy SF, Williams LT, Escobedo J;
XX DR WPI; 2003-577296/54.
XX PT Diagnosing prostate cancer, comprises detecting the presence or absence
XX PT of HML-2 (human endogenous retrovirus) expression product in a patient
XX PT sample.
XX PS Disclosure; SEQ ID NO 5; 117bp; English.
XX PS The present invention describes a method for diagnosing prostate cancer,
XX CC comprising detecting the presence or absence of HML-2 expression product
XX CC in a patient sample, where the expression product is produced by a
XX CC splicing event in which the 5' region and start codon of the env coding
XX CC region are joined to a downstream coding region in the reading frame +2
XX CC relative to that of env. Also described: (1) an isolated polynucleotide;
XX CC (2) an isolated polypeptide; (3) an antibody that binds to the
XX CC polypeptide; and (4) an immunogenic composition. Prostate cancer related
XX CC sequences have cytosolic, neuroprotective, antidiabetic and
XX CC immunostimulant activities, and can be used in vaccines and in gene
XX CC therapy. A polynucleotide, polypeptide or antibody of the invention can
XX CC be used in the manufacture of a medicament for preventing or treating
XX CC prostate, breast or testicular cancer, multiple sclerosis or insulin-
XX CC dependent diabetes mellitus. The present sequence represents a sequence
XX CC which is used in the exemplification of the present invention.
XX SQ Sequence 879 BP; 231 A; 212 C; 211 G; 225 T; 0 U; 0 Other;
XX Query Match 100.0%; Score 879; DB 9; Length 879;
XX Best Local Similarity 100.0%; Pred. No. 4.3e-288;
XX Matches 879; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGTGGGAAAAAGCAAGAGATGATGTTACTGTCGTCTGTAGAAAGAGATGACAT 60
DB 1 TGTGGGAAAAAGCAAGAGATGATGTTACTGTCGTCTGTAGAAAGAGATGACAT 60
QY 61 AGAGACTCCATTTTGTATGATCTAAGAAAAATTTCTTCGCTTGAAGATTTGTAATC 120
DB 61 AGAGACTCCATTTTGTATGATCTAAGAAAAATTTCTTCGCTTGAAGATTTGTAATC 120
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGCTGTCACTCAGGGTTAA 180
DB 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGCTGTCACTCAGGGTTAA 180
QY 181 ATGGATTAAAGGGGCGTGAGAGATGTCCTTGTAAACAGATGCTTGAAGGAGATGCTC 240
DB 181 ATGGATTAAAGGGGCGTGAGAGATGTCCTTGTAAACAGATGCTTGAAGGAGATGCTC 240
```

QY 241 CTTAAGAGTCATCAGCACTCCCTAATCTCAAGTACCAGGAGCAAAACCTGGGAAAG 300
DB 241 CTTAAGAGTCATCAGCACTCCCTAATCTCAAGTACCAGGAGCAAAACCTGGGAAAG 300
QY 301 CCGGAGGAGCCTCTGCTAGGAAAGCCAGGATTTGTCACAGTTCTCCCATGTGATAG 360
DB 301 CCGGAGGAGCCTCTGCTAGGAAAGCCAGGATTTGTCACAGTTCTCCCATGTGATAG 360
QY 361 CTTAAATATAGCCTCTGCTAGGAAAGCCAGGATTTGTCACAGTTCTCCCATGTGATAG 420
DB 361 CTTAAATATAGCCTCTGCTAGGAAAGCCAGGATTTGTCACAGTTCTCCCATGTGATAG 420
QY 421 AAGGCTCTGCTAGGAGATTTAGTAAAGAGGAAATGCTCTTGCAGTTGAGACA 480
DB 421 AAGGCTCTGCTAGGAGATTTAGTAAAGAGGAAATGCTCTTGCAGTTGAGACA 480
QY 481 AGAGGAAGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 481 AGAGGAAGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 541 GATTGTATGCTCCATCTTACTAGATAGGAAAGCCGCTTAAAGGCTGAGAGTGAACCT 600
DB 541 GATTGTATGCTCCATCTTACTAGATAGGAAAGCCGCTTAAAGGCTGAGAGTGAACCT 600
QY 601 GCGGAGGAGCATTCTGCTTGTAAAGCAGTGAATGTTATGATGATGATGATGATGATGAT 660
DB 601 GCGGAGGAGCATTCTGCTTGTAAAGCAGTGAATGTTATGATGATGATGATGATGATGAT 660
QY 661 CACAGCACTTAATCCTTATCAATGCTATGATGAGAAAGCCTTGTTCACATGTTTGTCT 720
DB 661 CACAGCACTTAATCCTTATCAATGCTATGATGAGAAAGCCTTGTTCACATGTTTGTCT 720
QY 721 GCTGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 721 GCTGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY 781 CCACAGATGATCAGTAAATCTAAGGAAAGCAGAGGCTGAGGAGTCTGATGATGCTG 840
DB 781 CCACAGATGATCAGTAAATCTAAGGAAAGCAGAGGCTGAGGAGTCTGATGATGCTG 840
QY 841 AACGCTGTTCCCGGCTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 879
DB 841 AACGCTGTTCCCGGCTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 879

RESULT 3
ABX04691
ID ABX04691 standard; DNA; 968 BP.
XX
AC ABX04691;
XX
DT 14-JAN-2003 (first entry)
XX
DE Human endogenous retrovirus k (herv-k) long terminal repeat #1.
XX
KM Human; endogenous retrovirus; herv; prostate cancer; testicular cancer;
KM multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
KM cancer; transgenic animal; de.
XX
OS Human endogenous retrovirus.
XX
PN NO20024647-A2.
XX
PD 13-JUN-2002.
XX
PF 07-DEC-2001; 2001WO-US047824.
XX
PR 07-DEC-2000; 2000US-0251830P.
XX
PR 07-DEC-2001; 2001US-00016504.
XX
PA (CHIR) CHIRON CORP.
XX

QY 241 CTTAAGAGTCATCAGCACTCCCTAATCTCAAGTACCAGGAGCAAAACCTGGGAAAG 300
DB 241 CTTAAGAGTCATCAGCACTCCCTAATCTCAAGTACCAGGAGCAAAACCTGGGAAAG 300
QY 301 CCGGAGGAGCCTCTGCTAGGAAAGCCAGGATTTGTCACAGTTCTCCCATGTGATAG 360
DB 301 CCGGAGGAGCCTCTGCTAGGAAAGCCAGGATTTGTCACAGTTCTCCCATGTGATAG 360
QY 361 CTTAAATATAGCCTCTGCTAGGAAAGCCAGGATTTGTCACAGTTCTCCCATGTGATAG 420
DB 361 CTTAAATATAGCCTCTGCTAGGAAAGCCAGGATTTGTCACAGTTCTCCCATGTGATAG 420
QY 421 AAGGCTCTGCTAGGAGATTTAGTAAAGAGGAAATGCTCTTGCAGTTGAGACA 480
DB 421 AAGGCTCTGCTAGGAGATTTAGTAAAGAGGAAATGCTCTTGCAGTTGAGACA 480
QY 481 AGAGGAAGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 481 AGAGGAAGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 541 GATTGTATGCTCCATCTTACTAGATAGGAAAGCCGCTTAAAGGCTGAGAGTGAACCT 600
DB 541 GATTGTATGCTCCATCTTACTAGATAGGAAAGCCGCTTAAAGGCTGAGAGTGAACCT 600
QY 601 GCGGAGGAGCATTCTGCTTGTAAAGCAGTGAATGTTATGATGATGATGATGATGATGAT 660
DB 601 GCGGAGGAGCATTCTGCTTGTAAAGCAGTGAATGTTATGATGATGATGATGATGATGAT 660
QY 661 CACAGCACTTAATCCTTATCAATGCTATGATGAGAAAGCCTTGTTCACATGTTTGTCT 720

PI Garcia P, Hardy SF, Williams LT, Escobedo J;
XX
XX WPI: 2002-691475/74.
XX
PT Novel isolated polypeptides useful for diagnosis of prostate cancer.
XX
PS Claim 18; Page 148; 152pp; English.
XX
CC The invention describes novel isolated polypeptides (I, Ib) useful for
CC diagnosing prostate cancer comprising obtaining a patient sample
CC containing prostate cells and detecting the presence or absence of an
CC expression product of a HML-2 endogenous retrovirus in a patient sample.
CC Polynucleotides associated with (I) are useful for diagnosis or treatment
CC of testicular cancer, multiple sclerosis or insulin-dependent diabetes
CC mellitus. An inhibitor of a HML-2 protease and a transdominant negative
CC mutant of HML-2 CORF are also useful in the manufacture of a medicament
CC for treating prostate cancer. (I) and (Ib) are useful for generating
CC antibodies specific to the polypeptides associated with cancer, as
CC targets for therapeutic intervention, and in immunising a transgenic
CC animal. This sequence represents a region of human endogenous retrovirus
CC (herv) of the HML-2 sub-group
XX
SQ Sequence 968 BP; 249 A; 242 C; 225 G; 252 T; 0 U; 0 Other;
Query Match 100.0%; Score 879; DB 6; Length 968;
Best Local Similarity 100.0%; Pred. No. 4,6e-288;
Matches 879; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGTGGGAAAAAGCAAGAGATCAGATTTGTTACTGTCTGTGTAGAAAGATGACAT 60
DB 1 TGTGGGAAAAAGCAAGAGATCAGATTTGTTACTGTCTGTGTAGAAAGATGACAT 60
QY 61 AGAGNACTCCATTTTGTATGATAGAAAGAAATTTCTTGTGCTGAGATTTCTTATC 120
DB 61 AGAGNACTCCATTTTGTATGATAGAAAGAAATTTCTTGTGCTGAGATTTCTTATC 120
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGCTGCTCACTCAGGTTAA 180
DB 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGCTGCTCACTCAGGTTAA 180
QY 181 ATGATTTAAGGCGGTGAGAGATGCTTTGTTAAAGATGCTTGAAGGAGCATGCTC 240
DB 181 ATGATTTAAGGCGGTGAGAGATGCTTTGTTAAAGATGCTTGAAGGAGCATGCTC 240
QY 241 CTTAAGAGTCATCAGCACTCCCTAATCTCAAGTACCAGGAGCAAAACCTGGGAAAG 300
DB 241 CTTAAGAGTCATCAGCACTCCCTAATCTCAAGTACCAGGAGCAAAACCTGGGAAAG 300
QY 301 CCGGAGGAGCCTGCTGCTAGGAAAGCCAGGATTTGTCACAGTTCTCCCATGTGATAG 360
DB 301 CCGGAGGAGCCTGCTGCTAGGAAAGCCAGGATTTGTCACAGTTCTCCCATGTGATAG 360
QY 361 CTTAAATATAGCCTCTGCTAGGAAAGCCAGGATTTGTCACAGTTCTCCCATGTGATAG 420
DB 361 CTTAAATATAGCCTCTGCTAGGAAAGCCAGGATTTGTCACAGTTCTCCCATGTGATAG 420
QY 421 AAGGCTCTGCTAGGAGATTTAGTAAAGAGGAAATGCTCTTGCAGTTGAGACA 480
DB 421 AAGGCTCTGCTAGGAGATTTAGTAAAGAGGAAATGCTCTTGCAGTTGAGACA 480
QY 481 AGAGGAAGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 481 AGAGGAAGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 541 GATTGTATGCTCCATCTTACTAGATAGGAAAGCCGCTTAAAGGCTGAGAGTGAACCT 600
DB 541 GATTGTATGCTCCATCTTACTAGATAGGAAAGCCGCTTAAAGGCTGAGAGTGAACCT 600
QY 601 GCGGAGGAGCATTCTGCTTGTAAAGCAGTGAATGTTATGATGATGATGATGATGATGAT 660
DB 601 GCGGAGGAGCATTCTGCTTGTAAAGCAGTGAATGTTATGATGATGATGATGATGATGAT 660
QY 661 CACAGCACTTAATCCTTATCAATGCTATGATGAGAAAGCCTTGTTCACATGTTTGTCT 720

DE Human reproductive system related antigen DNA SEQ ID NO: 8024.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 05-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.

PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251866P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.

PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
XX
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
XX PT used in preventing, treating or ameliorating a medical condition.
XX
XX
XX PS Disclosure; SEQ ID NO 8024; 1297bp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX CC number of human reproductive system related antigens. These can be used
XX CC in the prevention and treatment of reproductive system disorders,
XX CC including cancer. The present sequence is a genomic sequence encoding a
XX CC protein of the invention
XX
XX SQ Sequence 32249 BP; 10716 A; 6390 C; 5851 G; 9292 T; 0 U; 0 Other;

Query Match 97.5%; Score 856.6; DB 4; Length 32249;
Best Local Similarity 98.4%; Pred. No. 1.6e-279;
Matches 865; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTGCTGTGTAGAAAGAGATGACAT 60
Db 25971 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTGCTGTGTAGAAAGAGATGACAT 26030
QY 61 AGGAGACTCATTTTGTATGACTAAGAAAAATCTTGTGCTTGAATTCTGTATATC 120
Db 26031 AGGAGACTCATTTTGTATGACTAAGAAAAATCTTGTGCTTGAATTCTGTATATC 26090
QY 121 TATGACCTTATCCCAACCCCGTCTCTTGAACATGTGCTGTCTCATCAGGGTTAA 180
Db 26091 TATGACCTTATCCCAACCCCGTCTCTTGAACATGTGCTGTCTCATCAGGGTTGA 26150
QY 181 ATGGATTAAAGGCGGTGCAAGATGTCTTTGTTAAAGATGCTTGAAGGACATATCTC 26210
Db 26151 ATGGATTAAAGGCGGTGCAAGATGTCTTTGTTAAAGATGCTTGAAGGACATATCTC 26210
QY 241 CTTAAGAGTATCAACACTCCCTATATCTCAAGTACCCAGGAGACAAAACTGGGAAAG 300
Db 26211 CTTAAGAGTATCAACACTCCCTATATCTCAAGTACCCAGGAGACAAAACTGGGAAAG 26270
QY 301 CCGGAGGAGACTCTGCTTAGAAAGCAAGTATTTCCAAAGTTTCTCCCATGTGATAG 360
Db 26271 CCGGAGGAGACTCTGCTTAGAAAGCAAGTATTTCCAAAGTTTCTCCCATGTGATAG 26330
QY 361 CTTGAATATAGGCTCTGTTGGGAAAGGAAAGCTGACCGTCCCGACCCGAGACCCGTA 420
Db 26331 CTTGAATATAGGCTCTGTTGGGAAAGGAAAGCTGACCGTCCCGACCCGAGACCCGTA 26390
QY 421 AAGGCTGTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTTGGAGTTGAGACA 480
Db 26391 AAGGCTGTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTTGGAGTTGAGACA 26450
QY 481 AGAGAAAGGATCTGTCTCTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 26451 AGAGAAAGGATCTGTCTCTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26510
QY 541 GATTGTATGCTCATCTACTAGATAGAGTAAAGAAACCGCTTAGGGCTGAGGGTGGAGCT 600
Db 26511 GATTGTATGCTCATCTACTAGATAGAGTAAAGAAACCGCTTAGGGCTGAGGGTGGAGCT 26570
QY 601 GCGGCGACATATCTGCTTTGTAAAGCACTGAGATGTTTATGTGTATGATCTAAAG 660
Db 26571 GTGGGCGACATATCTGCTTTGTAAAGCACTGAGATGTTTATGTGTATGATCTAAAG 26630
QY 661 CACAGCCTTAATCTTTTACATTTGTATGATGCAAAAGCTTTGTTCAATGTTTGTCT 720
Db 720 CACAGCCTTAATCTTTTACATTTGTATGATGCAAAAGCTTTGTTCAATGTTTGTCT 720

Db 26631 CACAGCCTTAATCTTTTACATTTGTATGATGCAAAAGCTTTGTTCAATGTTTGTCT 26690
QY 721 GCTGACCTCTCCCAACAATTTGTTGTGACCTTGAACATCCCTCTTTCAGAAACAC 780
Db 26691 GCTGACCTCTCCCAACAATTTGTTGTGACCTTGAACATCCCTCTTTCAGAAACAC 26750
QY 781 CCACAGATGATCTGTTAAATCTAAGGAACTCAGAGGCTGGCGGGATCTTCCATATGCTG 840
Db 26751 CCACAGATGATCTGTTAAATCTAAGGAACTCAGAGGCTGGCGGGATCTTCCATATGCTG 26810
QY 841 AACGCTGTTCCCGGGATCCCTCTTCTTCTATATA 879
Db 26811 AACGCTGTTCCCGGGATCCCTCTTCTTCTATATA 26849
RESULT 6
ID ABL98205 standard; DNA; 32249 BP.
XX ABL98205;
XX
XX
XX 21-JUN-2002 (first entry)
DE Human testicular antigen encoding DNA fragment SEQ ID NO: 2857.
XX
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
XX KM reproductive system disorder; urinary system disorder; gene therapy;
XX KM cardiovascular disorder; respiratory disorder; neurological disorder;
XX KM gastrointestinal disease; infection; cytostatic; gene; ds.
XX
XX Homo sapiens.
XX
XX WO200155317-A2.
XX
XX 02-AUG-2001.
PD
PF 17-JAN-2001; 2001MO-US001329.
XX
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232337P.
PR 14-SEP-2000; 2000US-0232338P.
PR 14-SEP-2000; 2000US-0232339P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 25-SEP-2000; 2000US-0234999P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236328P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.

PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249219P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249248P.
PR 17-NOV-2000; 2000US-0249249P.
PR 17-NOV-2000; 2000US-0249250P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250319P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251858P.
PR 08-DEC-2000; 2000US-0251859P.
PR 08-DEC-2000; 2000US-0251860P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2001US-0259678P.
PR
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-483232/52.
XX
XX Nucleic acids encoding 973 human testicular antigen polypeptides, useful
PT for preventing, diagnosing and/or treating testicular cancer.
XX
XX Disclosure; SEQ ID NO 2857; 766pp; English.
XX
XX The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancer. The present sequence is a DNA encoding a
CC protein fragment of the invention
XX
SQ Sequence 32249 BP; 10716 A; 6390 C; 5851 G; 9292 T; 0 U; 0 Other;
Query Match 97.5%; Score 856.6; DB 4; Length 32249;
Best Local Similarity 98.4%; Pred. No. 1.ee-279;
Matches 865; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1 TGTGGGAAAAGAGAGAGATGATTGTTACTGTCTGTGTAGAAAAGATGACAT 60
Db '25971 TGTGGGAAAAGAGAGAGATGATTGTTACTGTCTGTGTAGAAAAGATGACAT 26030
QY 61 AGGAGACTTCATTTTGTATGTACTAGAAAATTTCTTGCCTTGAGATTCTTATC 120
Db 26031 AGGAGACTTCATTTTGTATGTACTAGAAAATTTCTTGCCTTGAGATTCTTATC 26090
QY 121 TATGACCTTACCCCAACCCCGCTCTGTGAACATGTCTGTCTCACTCAGGTTAA 180
Db 26091 TATGACCTTACCCCAACCCCGCTCTGTGAACATGTCTGTCTCACTCAGGTTGA 26150
QY 181 ATGATTTAAGGGGGGTCAGATGTGCTTGTAAACAGATGCTTGAAGGACATGCTC 240
Db 26151 ATGATTTAAGGGGGGTCAGATGTGCTTGTAAACAGATGCTTGAAGGACATGCTC 26210
QY 241 CTTAAGAGTCATCACTCCTTATCTCAAGTACCCAGGACACAAAATCTCGGAAGG 300
Db 26211 CTTAAGAGTCATCACTCCTTATCTCAAGTACCCAGGACACAAAATCTCGGAAGG 26270

OY	301	CCCGAGGAGACTCTGCGCTAGAGAAAACGACGGATATTGTCGCAACGTTCTCCCAATGAGATAG	360
Db	26271	CCCGAGGAGACTCTGCGCTAGAGAAAACGACGGATATTGTCGCAACGTTCTCCCAATGAGATAG	26350
OY	361	CCTGAATATATGAGCTCGTCGGGAAAGGGAAMAACCTGACCGTCCCGCCAGCCGACACCCGTA	420
Db	26331	TCTGAATATATGAGCTCGTCGGGAAAGGGAAMAACCTGACCGTCCCGCCAGCCGACACCCGTA	26390
OY	421	AAGGCTCTGTCGTCGAGGAGATTGTAATAAGAGAAAGAAATGCTCTTGACGTTGAGACA	480
Db	26391	AAGGCTCTGTCGTCGAGGAGATTGTAATAAGAGAAAGAAATGCTCTTGACGTTGAGACA	26450
OY	481	AGAGAAAGGACATCTGTCTCTGCTGCTGCTCGGCGAATGGAATGTCCTGGTATATAAAACC	540
Db	26451	AGAGAAAGGACATCTGTCTCTGCTGCTGCTCGGCGAATGGAATGTCCTGGTATATAAAACC	26510
OY	541	GATTGTATGCTCCATCTACTGAGATAGGAGAAAAACGCGCTTAGGCTGAGAGTGGGACT	600
Db	26511	GATTGTATGCTCCATCTACTGAGATAGGAGAAAAACGCGCTTAGGCTGAGAGTGGGACT	26570
OY	601	GCGGGACGCAATACCTGCTTTGTAAAGCATGAGATGTTATGTGATGCAATCTTAAAG	660
Db	26571	GTGGGACGCAATACCTGCTTTGTAAAGCATGAGATGTTATGTGATGCAATCTTAAAG	26630
OY	661	CACAGCACTTAATCTCTTACATGTCTATGATGCAAAAGACTTGTTCACATGTTTGTCT	720
Db	26631	CACAGCACTTAATCTCTTACATGTCTATGATGCAAAAGACTTGTTCACATGTTTGTCT	26690
OY	721	GCTAACCCCTCCCCACAAATTGTCTTGTGACCTGACACATCCCGCTCTCGAGAAACAC	780
Db	26691	GCTAACCCCTCCCCACAAATTGTCTTGTGACCTGACACATCCCGCTCTCGAGAAACAC	26750
OY	781	CCACAGATGATCAGTAATACTAAAGGAACTCAGAAGCTGGCGGATCTCCATATGCTG	840
Db	26751	CGACAGATGATCAGTAATACTAAAGGAACTCAGAAGCTGGCGGATCTCCATATGCTG	26810
OY	841	AACGCTGTTCCCCGGGTCCCTCTCTTTCTCTATA	879
Db	26811	AACGCTGTTCCCCGGGTCCCTCTCTTTCTCTATA	26849

RESULT 7	
ABX04612	
ID	ABX04612 standard; DNA; 9179 BP.
XX	
AC	ABX04612;
XX	
DT	14-JAN-2003 (first entry)
XX	
DE	Human endogenous retrovirus k10 (hervk10) polynucleotide.
XX	
KW	Human; endogenous retrovirus; henv; prostate cancer; testicular cancer;
KW	multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
KW	cancer; transgenic animal; ds.
XX	
OS	Human endogenous retrovirus.
XX	
PN	WO200246477-A2.
XX	
PD	13-JUN-2002.
XX	
PF	07-DEC-2001; 2001WO-US047824.
XX	
PR	07-DEC-2000; 2000US-0251830P.
PR	07-DEC-2001; 2001US-00016604.
XX	
PA	(CHIR) CHIRON CORP.
XX	
PI	Garcia P, Hardy SF, Williams LT, Escobedo J;
XX	
DR	WPI; 2002-691475/74.
XX	

Novel isolated polypeptides useful for diagnosis of prostate cancer
Disclosure; Page 137-138; 152pp; English.

The invention describes novel isolated polypeptides (I, Ib) useful for diagnosing prostate cancer comprising obtaining a patient sample containing prostate cells and detecting the presence or absence of an expression product of a HmL-2 endogenous retrovirus in a patient sample. Polynucleotides associated with (I) are useful for diagnosis or treatment of testicular cancer, multiple sclerosis or insulin-dependent diabetes mellitus. An inhibitor of a HmL-2 protease and a transdominant negative mutant of HmL-2 CORF are also useful in the manufacture of a medicament for treating prostate cancer. (I) and (Ib) are useful for generating antibodies specific to the polypeptides associated with cancer, as targets for therapeutic intervention, and in immunising a transgenic animal. This sequence represents a polynucleotide from the human endogenous retrovirus (herv) of the HmL-2 sub-group

Sequence 9179 BP; 2959 A; 1866 C; 1952 G; 2402 T; 0 U; 0 Other;

Query Match 97.3%; Score 855; DB 6; Length 9179;

```

Best Local Similarity      20.58;      E-Val: 2.00e-272
Matches      864; Conservative      0; Mismatches      15; Indels      0; Gaps      0;

```

Qy 1 TGTGGGGAAGCAAGAGATCAGATTGTTACTGTCTGTCTGTGTGAAAGAGTAGACAT 60

Db 1 TGTGGGGAAGCAAGAGATCAATTGTTACTGTGTCTGTGTGAAAGAGTAGACAT 60

61 AGGAGACTTCATTCTTGTATGTACTAAGAAATAATCTCTGCCCTTGAGATTCTGTTAATC 120

61 AGGAGACATTCATTTGTTATGTGCTAAGAAAAATTCCTCTGCCCCGAGATTCTGTATC 120

121 TATGACCTTACCCCAACCCGTCCTCTGAAACATGTCCTGTCACACTCAGGGTTAA 180

[illegible]

101 240

240

QUESTION

1. *What is the purpose of this study?*

[illegible][illegible]

UD 381 1C1GHHH1H1GGL1CG1GGGHHGGGHHHGHAC1GHACCG1CCCCCGHCCCCGGHACCC1G1H 320

421 AAGGCTGTGCTGAGGAGATAGTAAAGAGGAAAGGAAGCCCTCTGCAATGAGACA 480

Db 421 AAGGTCCTGCTGAGGAGATAGTAAAGAGGAAGGATGCCCTCTGCAGTTCAGACA 480

481 AGAGGAAGGCATCTGTCTCTCCCTGTCCCTGGGCATGGAATGTCGGTATAAAACC 540

Db 481 AGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCC 540

541 GATTGTATGCTCCATCTACTGAGATAGGGAACCGCCTTAGGGCTGAGGTTGGACCT 600

Db 541 GATTGATGCTCCATCTACTGAGATAGGGAACCGCCTTAGGGCTGAGGTGGACCT 600

601 GCGGCAGCACTGCTTTGTAAGCACTGAGATGTTATGTGATGCATATCTAAAG 660

Db 601 GCGGCAGCACTGCTTGTAAAGCATTGAGATGTTATGTGTATGCATATCCAAG 660

661 CACAGCACTTAATCCTTTACATTGTCTATGATGCAAGACCTTGTTCACATGTTGTCT 720

Db 661 CACAGCACTTAATCCTTACATTGTCATATGATGCCAAGACCTTTGTTTCACGTGTTGCT 720

721 GCTGACCCCTCTCCCCACAAATTGTTGTGACCCCTGACACATCCCCCTTTGAGAAACAC 780

PD 19-JUN-2003.
XX
XX 09-DEC-2002; 2002MO-US039344.
XX
XX 07-DEC-2001; 2001US-00016604.
PR 07-DEC-2001; 2001US-0340064P.
PR 07-DEC-2001; 2001MO-US047824.
PR 12-JUN-2002; 2002US-0398046P.
XX
XX (CHIR) CHIRON CORP.
PI Garcia P, Hardy SF, Williams LT, Escobedo J;
XX WPI; 2003-577296/54.
XX
XX Diagnosing prostate cancer, comprises detecting the presence or absence
PT of HML-2 (human endogenous retrovirus) expression product in a patient
PT sample.
PS Disclosure; SEQ ID NO 46; 117bp; English.
XX
XX The present invention describes a method for diagnosing prostate cancer,
CC comprising detecting the presence or absence of HML-2 expression product
CC in a patient sample, where the expression product is produced by a
CC splicing event in which the 5' region and start codon of the env coding
CC region are joined to a downstream coding region in the reading frame +2
CC relative to that of env. Also described: (1) an isolated polynucleotide;
CC (2) an isolated polypeptide; (3) an antibody that binds to the
CC polypeptide; and (4) an immunogenic composition. Prostate cancer related
CC sequences have cytostatic, neuroprotective, antidiabetic and
CC immunostimulant activities, and can be used in vaccines and in gene
CC therapy. A polynucleotide, polypeptide or antibody of the invention can
CC be used in the manufacture of a medicament for preventing or treating
CC prostate, breast or testicular cancer, multiple sclerosis or insulin-
CC dependent diabetes mellitus. The present sequence represents a sequence
CC which is used in the exemplification of the present invention.
XX
XX
SQ Sequence 968 BP; 248 A; 238 C; 226 G; 256 T; 0 U; 0 Other;
Query Match 96.9%; Score 851.8; DB 9; Length 968;
Best Local Similarity 98.1%; Pred. No. 8.4e-279;
Matches 862; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 1 TGTGGGAAAAGCAAGAGATCAGATTGTCTGTCTGTGTAGAAAGATGACAT 60
DB 1 TGTGGGAAAAGCAAGAGATCAGATTGTCTGTCTGTGTAGAAAGATGACAT 60
QY 61 AGGAGATCTCCATTTTGTATGTACTAAGAAAATTTCTTGCCTTGAGATTCTGTATC 120
DB 61 GGGAGACTCCATTTTGTATGTCTAAGAAAATTTCTTGCCTTGAGATTCTGTATC 120
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTCACTCAGGGTTAA 180
DB 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTCACTCAGGGTTGA 180
QY 181 ATGGATTAAAGGCGGTGAGAGATGCTTTGTAAAGATGCTTGAAGGAGCATGCTC 240
DB 181 ATGGATTAAAGGCGGTGAGAGATGCTTTGTAAAGATGCTTGAAGGAGCATGCTC 240
QY 241 CTTTAAAGTCAACCACTCCCTATATCTCAAGTACCAAGGACAAAATCTGGGAAG 300
DB 241 CTTTAAAGTCAACCACTCCCTATATCTCAAGTACCAAGGACAAAATCTGGGAAG 300
QY 301 CCGGAGGAGCTGCTGCTAGGAAAGCAGATGTGCAAGTTCCTCCCATGTGATAG 360
DB 301 CCGGAGGAGCTGCTGCTAGGAAAGCAGATGTGCAAGTTCCTCCCATGTGATAG 360
QY 361 CTTGAAATATGCTCGTGGGAAAGAAAGATGACGTCCTCCGACCCGACCCGTA 420
DB 361 TCTGAAATATGCTCGTGGGAAAGAAAGATGACGTCCTCCGACCCGACCCGTA 420
QY 421 AAGGCTGTGCTGAGAGATTATTAAGAGAGAAATGCTCTTTCAGATTGAGACA 480
DB 421 AAGGCTGTGCTGAGAGATTATTAAGAGAGAAATGCTCTTTCAGATTGAGACA 480

DB 421 AAGGCTGTGCTGAGAGATTATTAAGAGAGAAATGCTCTTTCAGATTGAGACA 480
QY 481 AGAGAAAGCATGTCTCTGCTGCTGCTGCGGCAATGATGTCTGATATAAAACC 540
DB 481 AGAGAAAGCATGTCTCTGCTGCTGCTGCGGCAATGATGTCTGATATAAAACC 540
QY 541 GATTGTATGTCTCATCTAAGATGAGGAAAAACCGCTTACGGCTGAGGCTGACCT 600
DB 541 GATTGTATGTCTCATCTAAGATGAGGAAAAACCGCTTACGGCTGAGGCTGACCT 600
QY 601 GCGGAGAGCAATCTGCTTTGTAAGACATGATGTTATGTATGATATCTAAAG 660
DB 601 GCGGAGAGCAATCTGCTTTGTAAGACATGATGTTATGTATGATATCTAAAG 660
QY 661 CACAGACCTTAATCTTTACATTTGTATGATGATCAAGACCTTTGTACATGTTGCT 720
DB 661 CACAGACCTTAATCTTTACATTTGTATGATGATCAAGACCTTTGTACATGTTGCT 720
QY 721 GCTGACCTCTCCCAACATTTGTTGTGACCTGACATCCCTCTTTCGAGAAACAC 780
DB 721 GCTGACCTCTCCCAACATTTGTTGTGACCTGACATCCCTCTTTCGAGAAACAC 780
QY 781 CCAAGATGATCAGTAATTAATTAAGGAACTGAGAGCTGGGGATCCCTCATATGCTG 840
DB 781 CCAAGATGATCAGTAATTAATTAAGGAACTGAGAGCTGGGGATCCCTCATATGCTG 840
QY 841 AAGCGTGTTCCTCCGAGTCCCTCTTCTTCTATATA 879
DB 841 AAGCGTGTTCCTCCGAGTCCCTCTTCTTCTATATA 879
RESULT 10
AAD00147
ID AAD00147 standard; DNA; 35414 BP.
XX
AC AAD00147;
XX
XX 31-JUL-2000 (first entry)
XX
DE TR12 related DNA-11.
XX
XX Human tumour necrosis factor receptor related gene 12; TR12; cytostatic;
KW anti-infective; anti-human immunodeficiency virus; nephrotoxic; cancer;
KW anti-arteriosclerotic; analgesic; cardiact; immunosuppressant; virucide;
KW anti-allergic; anti-inflammatory; hepatotropic; antidiabetic; vulnery;
KW anti-asthmatic; antiparasitic; antibacterial; antitumor; angiogenic;
KW microbial infection; nephritis; bone disease; pain; atherosclerosis;
KW cardiovascular disorder; allergy; inflammation; asthma; cardiomyopathy;
KW neurodegenerative disease; Alzheimer's disease; graft rejection;
KW liver disease; autoimmune disease; diabetes; influenza; psoriasis;
KW glomerulonephritis; septic shock; ulcerative colitis; angiogenesis;
KW hematoptosis; wound healing; ds.
XX
OS Homo sapiens.
XX
PN WO200023572-A1.
XX
PD 27-APR-2000.
XX
XX 19-OCT-1999; 99MO-US024413.
XX
XX 20-OCT-1998; 98US-0104950P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ni J, Ruben SM,
XX
XX WPI; 2000-339675/29.
XX
XX Tumor necrosis factor receptor related gene 12 (TR12) polynucleotides and
PT polypeptides, useful for prevention, treatment, and diagnosis of various
PT diseases such as cancer, atherosclerosis, pain, and cardiovascular
PT disorders.

XX	
PS	Disclosure; Page 227-246; 250pp; English.
XX	

CC The patent discloses the use of tumour necrosis factor receptor related
CC gene 12 (TR12). TR12 gene and protein are useful for treatment of
CC diseases including, cancer, microbial infections (including human
CC immunodeficiency virus (HIV)), nephritis, bone diseases, atherosclerosis,
CC pain, cardiovascular disorders, acquired immune deficiency syndrome
CC (AIDS), allergy, inflammation, neurodegenerative diseases (e.g.
CC Alzheimer's, Parkinson's), graft rejection, liver disease, autoimmune
CC disease, cardiomyopathy, diabetes, influenza, asthma, psoriasis,
CC glomerulonephritis, septic shock and ulcerative colitis. They may also be
CC useful for promoting angiogenesis, regulating haematopoiesis, wound
CC healing, and has various diagnostic and research applications. The
CC present sequence is a specifically related polynucleotide sequence that
CC is specifically excluded from the scope of the present invention
XX
50 Sequence 35414 BP; 8874 A; 8376 C; 8358 G; 9806 T; 0 U; 0 Other;

Query Match	96.2%	Score 845.4	DB 3	Length 35414
Best Local Similarity	97.6%	Pred. NO. 1.1e-275		
Matches 858; Conservative	0	Mismatches 21	Indels 0	Gaps 0

QY	1	TGTGGGGAAAAGCAAGAGATACAAATTGTTACTGTGTCTGTGTGAAGAAGTGAACAT	60
Db	8681	TGTGGGGAAAAGCAAGAGATACAAATTGTTACTGTGTCTGTGTGAAGAAGTGAACAT	8746
QY	61	AGGAGACTCCATTTTGTATGTACTAAGAAAAATTCCTCTGCTTGAGATTCTGTTAATC	120
Db	8741	AGGAGACTCCATTTTGTATGTACTAAGAAAAATTCCTCTGCTTGAGATTCTGTTAATC	8800
QY	121	TATAGACCTTACCCCAACCCCGTGTCTGTAAACATGTGCTGTGTCCACTCAGGGTTAA	180
Db	8801	TATAGACCTTACCCCAACCCCGTGTCTGTAAACATGTGCTGTGTCCACTCAGGGTTGA	8860
QY	181	ATGATTTAAGGGCGGTGCAGATGTGCTTTGTAAACAGATGCTTGAAGCAGCATGCTC	240
Db	8861	ATGATTTAAGGGCGGTGCAGATGTGCTTTGTAAACAGATGCTTGAAGCAGCAGCTC	8922
QY	241	CTTTAAGACTATCACCACTCCCTTAATCTCAAGTACCCAGGACACAAAACTGCGGAAG	300
Db	8921	CTTTAAGACTATCACCACTCCCTTAATCTCAAGTACCCAGGACACAAAACTGCGGAAG	8980
QY	301	CCGAGAGGAGCCTCTGCGCTTAGGAAAAGCAGATATGTCCAAAGTTCTCCCATGTGATAG	360
Db	8981	CCGAGAGGACTCTGCGCTTAGGAAAAGCAGATATGTCCAAAGTTCTCCCATGTGATAG	9040
QY	361	CCTGAATATGCGCCTCTGTGGGAAGGAAAGACCTGACCCGACGCCAGCACCCCTGA	420
Db	9041	TCTGAATATGCGCCTCTGTGGGAAGGAAAGACCTGACCCGACGCCAGCACCTGTA	9100
QY	421	AAGGCTCTGTCTTAGAGAGATTAGTAAAGAGAGAGATGCTCTTGCACTGTGAGACA	480
Db	9101	AAGGCTCTGTCTTAGAGAGATTAGTAAAGAGAGAGATGCTCTTGCACTGTGAGACA	9160
QY	481	AGAGAGAGGACTGTGTCTCTGCGCTGCTCGGGCAATGGAATGCTCGGATATAAAACC	540
Db	9161	AGAGAGAGGACTGTGTCTCTGCGCTGCTCGGGCAATGGAATGCTCGGATATAAAACC	9220
QY	541	GATTGTATGCTCATCTAAGTAGATAGGAAAAAACCGCTTAGAGGCTGGAAGTGGGACT	600
Db	9221	GATTGTATGCTCATCTAAGTAGATAGGAAAAAACCGCTTAGAGGCTGGAAGTGGGACT	9280
QY	601	GGGGGAGAGCAATACTGCTTTGTAAAGCACTGAGATGTTATGTGTATGCAATATCTAAAG	660
Db	9281	GGGGGAGAGCAATACTGCTTTGTAAAGCACTGAGATGTTATGTGTATGCAATATCTAAAG	9340
QY	661	CACAGCACTTATCCTTACATTTGTCTATGTAGTGAAGAAGCCTTTGTTCACATGTTGTCT	720
Db	9341	CACAGCACTTATCCTTACATTTGTCTATGTAGTGAAGAAGCCTTTGTTCACATGTTGTCT	9400
QY	721	GCTGACCTCTCCCAACAATTGTCTGTGACCTTGACACATCCCTCTTGGAAGAACAC	780

Db	9401	GCTGACCTCTCCCAACAATATGTCITGAGCCCTGACACATCCCTCTTTGAGAAACAC	9462
Qy	781	CCACGATGATCAGTAAATACTAAGGGAACCTCAGAGCTGGCGGATCTCCATATGCTG	840
Db	9461	CCACGATGATGAAATTAATATCTAAGGGAACCTCAGAGCTGGCGGATCTCCATATGCTA	9520
Qy	841	AACGTGGTTCGCCGGGTCGCCCTCTTCTTCTCTATA	879
Db	9521	AACGTGGTTCGCCGGGTCGCCCTCTTATTTCTTCTATA	9559
RESULT 11			
ID	ABK84349/C	standard; cDNA: 222930 BP.	
AC	ABK84349;		
XX			
DT	14-AUG-2002	(first entry)	
DE	Human cDNA differentially expressed in granulocytic cells #920.		
XX			
KW	Human; ss; granulocytic cell; DNA chip; bacterial infection;		
KW	viral infection; parasitic infection; protozoal infection;		
KW	fungal infection; sterile inflammatory disease; psoriasis;		
KW	rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;		
KW	adult respiratory distress syndrome; inflammatory bowel disease;		
KW	Crohn's disease; ulcerative colitis; peridontal disease;		
XX	granulocyte activation; chronic inflammation; allergy.		
OS	Homo sapiens.		
PN	WO200228999-A2.		
XX			
PD	11-APR-2002.		
XX			
PF	03-OCT-2001; 2001WO-US030821.		
PR	03-OCT-2000; 2000US-0237189P.		
PA	(GENE-) GENE LOGIC INC.		
XX			
P1	Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;		
DR	WPI; 2002-435328/46.		
PT	Detecting granulocyte activation by detecting differential expression of		
PT	genes associated with granulocyte activation, which serves as diagnostic		
PT	markers that is useful for monitoring disease states and drug toxicity.		
XX	Claim 1; SEQ ID NO 920; 114pp; English.		
XX			
CC	The invention relates to detecting (M1) granulocyte (GC) activation		
CC	(GCA), by detecting the level of expression of gene(s) (Gs) identified by		
CC	DNA chip analysis as given in the specification, and comparing the		
CC	expression level to an expression level in an unactivated GC, where		
CC	differential expression of Gs is indicative of GCA. Also included are		
CC	modulating (M2) GA by contacting GC with an agent that alters the		
CC	expression of at least one gene in Gs; (2) screening (M3) for an agent		
CC	capable of modulating GCA or an inflammation (especially chronic) in a		
CC	tissue, an allergic response in a subject, exposure of a subject to a		
CC	pathogen or sterile inflammatory disease using the gene expression		
CC	profile; (3) detecting (M4) an inflammation (especially chronic) in a		
CC	tissue, an allergic response in a subject, exposure of a subject to a		
CC	pathogen or sterile inflammatory disease, by detecting the level of		
CC	expression in a sample of the tissue of gene(s) from Gs, where the level		
CC	of expression of the gene is indicative of inflammation; (4) treating		
CC	(M5) an inflammation (especially chronic) or in a tissue, an allergic		
CC	response in a subject, exposure of a subject to a pathogen or sterile		
CC	inflammatory disease, by contacting a tissue having inflammation with an		
CC	agent that modulates the expression of gene(s) from Gs in the tissue. M1		
CC	is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful		
CC	for screening an agent capable of modulating GCA preferably in an		

CC inflammation in a tissue; M4 is useful for detecting an inflammation
CC (especially chronic) in a tissue, an allergic response in a subject,
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC cardiac reperfusion injury, renal reperfusion injury, AIDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease, also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
CC conditions. The present sequence represents a gene differentially
CC expressed in granulocytes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
CC
CC
SQ Sequence 222930 BP; 53286 A; 56778 C; 57221 G; 55643 T; 0 U; 2 Other;

Query Match 95.6%; Score 840.6; DB 6; Length 222930;
Best Local Similarity 97.3%; Pred. No. 1.4e-273;
Matches 85; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
QY 1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTGCTGTGTGAGAAAGATGACAT 60
DB 98831 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTGCTGTGTGAGAAAGATGACAT 98772
QY 61 AGGAGACTCCATTTTGTATGTACTAAGAAAATTTCTGCTTGAGATTCTTTATC 120
DB 98771 AGGAGACTCCATTTTGTATGTACTAAGAAAATTTCTGCTTGAGATTCTTTATC 98712
QY 121 TATGACCTTACCCCAACCCCGTCTCTGTAACATGTGCTGTGCTCACTCAGGTTAA 180
DB 98711 TATGACCTTACCCCAACCCCGTCTCTGTAACATGTGCTGTGCTCACTCAGGTTAA 98652
QY 181 ATGATTTAAGGCGGTGAGATGTGCTTGTAAACAGATGCTGAAGGAGATGCTC 240
DB 98651 ATGATTTAAGGCGGTGAGATGTGCTTGTAAACAGATGCTGAAGGAGATGCTC 98592
QY 241 CTTAAGAGTCATCACCACCTCCCTATCTCAAGTACCCGAGACAAAATCTGGGAAAG 300
DB 98591 CTTAAGAGTCATCACCACCTCCCTATCTCAAGTACCCGAGACAAAATCTGGGAAAG 98532
QY 301 CCGGAGGACCTCTGCTAGGAAAAGCAGATTTGTTCCACGTTTCTCCCATGTATAG 360
DB 98531 CCGGAGGACCTCTGCTAGGAAAAGCAGATTTGTTCCACGTTTCTCCCATGTATAG 98472
QY 361 CTTGAATATATGCTCTGTGGAAGGAAAGCCTGACCGTCCCAAGCCGACACCCGTA 420
DB 98471 TCTGAATATATGCTCTGTGGAAGGAAAGCCTGACCGTCCCAAGCCGACACCCGTA 98412
QY 421 AAGGATCTGTGCTGAGAGATTTAGTAAAGAGAAAGATGCTCTTGCACTTGAGACA 480
DB 98411 AAGGATCTGTGCTGAGAGATTTAGTAAAGAGAAAGATGCTCTTGCACTTGAGACA 98352
QY 481 AGAGAAAGCATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 98351 AGAGAAAGCATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 98292
QY 541 GATTGTATGCTTCATCTACTAGATAGGAAAAACCGCTTAAGGCTGAGAGTGGACCT 600
DB 98291 GATTGTATGCTTCATCTACTAGATAGGAAAAACCGCTTAAGGCTGAGAGTGGACCT 98232
QY 601 GCGGGACGACATACCTGCTTTGTAAGCATGAGATGTTATGTATGATCATATCTAAAG 660
DB 98231 GCGGGACGACATACCTGCTTTGTAAGCATGAGATGTTATGTATGATCATATCTAAAG 98172
QY 661 CACAGACTTAATCTTTTACATTTGTCTATGATGCAAGACCTTTGTTCATGTTTGTCT 720
DB 98171 CACAGACTTAATCTTTTACATTTGTCTATGATGCAAGACCTTTGTTCATGTTTGTCT 98112
QY 721 GCTGACCTCTTCCCAACATTTGTCTGTGACCTTGACACATCCCTCTTGAAGAACAC 780
DB 98111 GCTGACCTCTTCCCAACATTTGTCTGTGACCTTGACACATCCCTCTTGAAGAACAC 98052
```

```
QY 781 CCAAGATGATGATTAATTAATTAAGGAATCAGAGGCTGGCCGATCTTCAATATGCTG 840
DB 98051 CCAAGATGATTAATTAATTAAGGAATCAGAGGCTGGCCGATCTTCAATATGCTG 97992
QY 841 AACGCTGTTCCTCCGGGTCCTTCTTCTTCTCTATA 879
DB 97991 AACGCTGTTCCTCCGGGTCCTTCTTCTTCTCTATA 97953
```

RESULT 12

ABX04695
ID ABX04695 standard; DNA: 968 BP.

AC ABX04695;

DT 14-JAN-2003 (first entry)

DE Human endogenous retrovirus k (herv-k) long terminal repeat #5.

KW Human; endogenous retrovirus; herv; prostate cancer; testicular cancer;
multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;

KM cancer; transgenic animal; de.

OS Human endogenous retrovirus.

PN WO200246477-A2.

PD 13-JUN-2002.

PF 07-DEC-2001; 2001MO-US047824.

PR 07-DEC-2000; 2000US-0251830P.

PR 07-DEC-2001; 2001US-00016604.

PA (CHIR) CHIRON CORP.

P1 Garcia P, Hardy SF, Williams LT, Escobedo J;

XX WPT, 2002-691475/74.

PT Novel isolated polypeptides useful for diagnosis of prostate cancer.

XX Claim 18; Page 149; 152pp; English.

CC The invention describes novel isolated polypeptides (I, Ib) useful for

CC diagnosing prostate cancer comprising obtaining a patient sample

CC containing prostate cells and detecting the presence or absence of an

CC expression product of a HML-2 endogenous retrovirus in a patient sample.

CC Polynucleotides associated with (I) are useful for diagnosis or treatme

CC of testicular cancer, multiple sclerosis or insulin-dependent diabetes

CC mellitus. An inhibitor of a HML-2 protease and a transdominant negative

CC mutant of HML-2 CORF are also useful in the manufacture of a medicament

CC for treating prostate cancer. (I) and (Ib) are useful for generating

CC antibodies specific to the polypeptides associated with cancer, as

CC targets for therapeutic intervention, and in immunising a transgenic

CC animal. This sequence represents a region of human endogenous retrovirus

CC (herv) of the HML-2 sub-group

XX
SQ Sequence 968 BP; 250 A; 237 C; 225 G; 256 T; 0 U; 0 Other;

Query Match 95.3%; Score 837.4; DB 6; Length 968;
Best Local Similarity 97.0%; Pred. No. 6.8e-274;
Matches 85; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```
QY 1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTGCTGTGTGAGAAAGATGACAT 60
DB 1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTGCTGTGTGAGAAAGATGACAT 60
QY 61 AGGAGACTCCATTTTGTATGTACTAAGAAAATTTCTGCTTGAGATTCTTTATC 120
DB 61 AGGAGACTCCATTTTGTATGTACTAAGAAAATTTCTGCTTGAGATTCTTTATC 120
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGCTGTGCTCACTCAGGTTAA 180
```

```
Db 121 TATTAACCTTACCCCAACCCCGTCTCTGAAACATGTCTATGTCAACTCAGAGTTGA 180
Qy 181 ATGATTAAGGCGGTCAGAGATGTCTTTTAAACAGATGCTTGAAGGAGCATGTCTC 240
Db 181 ATGATTAAGGCGGTCAGAGATGTCTTTTAAACAGATGCTTGAAGGAGCATGTCTC 240
Qy 241 CTTAAGAGTCATCAACATCCCTATCTCAAGTACCCAGGACACAAAACCTGGGAAAG 300
Db 241 CTTAAGAGTCATCAACATCCCTATCTCAAGTACCCAGGACACAAAACCTGGGAAAG 300
Qy 301 CCGGAGGACCTCTGCTAGGAAAGCCAGATATTGTCCAGTTTCTCCCATGTGATAG 360
Db 301 CCGGAGGACCTCTGCTAGGAAAGCCAGATATTGTCCAGTTTCTCCCATGTGATAG 360
Qy 361 CTGGAATATGAGCTCTGGGAAAGAAACCTGACCGTCCCGACCCGACACCGCTA 420
Db 361 CTGGAATATGAGCTCTGGGAAAGAAACCTGACCGTCCCGACCCGACACCGCTA 420
Qy 421 AAGGCTCTGTGCTAGAGAGATTAAGTAAAGAGAGAGATGCTTGGAGTGAACA 480
Db 421 AAGGCTCTGTGCTAGAGAGATTAAGTAAAGAGAGAGATGCTTGGAGTGAACA 480
Qy 481 AGAGAAAGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 481 AGAGAAAGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Qy 541 GATTGTATGCTCCATCTTAAGTAAAGGAAAGAAACCCGCTTGAAGGAGTGGACCT 600
Db 541 GATTGTATGCTCCATCTTAAGTAAAGGAAAGAAACCCGCTTGAAGGAGTGGACCT 600
Qy 601 GCGGGGACGACATATGCTTGTAAAGCAGTGAATGTTATGATGATGATGATGATGATG 660
Db 601 GCGGGGACGACATATGCTTGTAAAGCAGTGAATGTTATGATGATGATGATGATGATG 660
Qy 661 CACAGACCTTAATCTTATCTTATCTATGATGAAAGACCTTGTTCACATGTTTGTCT 720
Db 661 CACAGACCTTAATCTTATCTTATCTATGATGAAAGACCTTGTTCACATGTTTGTCT 720
Qy 721 GCTGACCTCTCCCAACATGCTTGTGACCTGACACATCCCTCTTGAAGAACAC 780
Db 721 GCTGACCTCTCTCCCAACATGCTTGTGACCTGACACATCCCTCTTGAAGAACAC 780
Qy 781 CCAAGATGATCACTAATACTAAGGAACTCAGAGGCTGGGAGATCTCCATATGCTG 840
Db 781 CCAAGATGATCACTAATACTAAGGAACTCAGAGGCTGGGAGATCTCCATATGCTG 840
Qy 841 AAGGCTGTTCCCGGCTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 879
Db 841 AAGGCTGTTCCCGGCTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 879

RESULT 13
ABX04694
ID ABX04694 standard; DNA; 968 BP.
AC ABX04694;
XX
XX 14-JAN-2003 (first entry)
XX
XX Human endogenous retrovirus k (herv-k) long terminal repeat #4.
XX
XX Human endogenous retrovirus; herv; prostate cancer; testicular cancer;
XX multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
XX cancer; transgenic animal; ds.
XX
XX Human endogenous retrovirus.
XX
XX WO20024647-A2.
XX
XX 13-JUN-2002.
XX
XX 07-DEC-2001; 2001WO-US047824.
XX
```

```
XX 07-DEC-2000; 2000US-0251830P.
PR 07-DEC-2001; 2001US-00016604.
PA (CHIR) CHIRON CORP.
PI Garcia P, Hardy SF, Williams LT, Escobedo J;
XX WPI; 2002-691475/74.
XX
XX Novel isolated polypeptides useful for diagnosis of prostate cancer.
XX Claim 18; Page 149; 152pp; English.
XX
XX The invention describes novel isolated polypeptides (I, Ib) useful for
XX diagnosing prostate cancer comprising obtaining a patient sample
XX containing prostate cells and detecting the presence or absence of an
XX expression product of a HML-2 endogenous retrovirus in a patient sample.
XX Polynucleotides associated with (I) are useful for diagnosis or treatment
XX of testicular cancer, multiple sclerosis or insulin-dependent diabetes
XX mellitus. An inhibitor of a HML-2 protease and a transdominant negative
XX mutant of HML-2 CORP are also useful in the manufacture of a medicament
XX for treating prostate cancer. (I) and (Ib) are useful for generating
XX antibodies specific to the polypeptides associated with cancer, as
XX targets for therapeutic intervention, and in immunising a transgenic
XX animal. This sequence represents a region of human endogenous retrovirus
XX (herv) of the HML-2 sub-group
XX
XX Sequence 968 BP; 252 A; 238 C; 224 G; 254 T; 0 U; 0 Other;
XX
XX Query Match 95.3%; Score 837.4; DB 6; Length 968;
XX Best Local Similarity 97.0%; Pred. No. 6.8e-274;
XX Matches 853; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
XX
Qy 1 TGTGGGGAAGAAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 60
Db 1 TGTGGGGAAGAAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 60
Qy 61 AGAGAGCTCCATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 120
Db 61 AGAGAGCTCCATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 120
Qy 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGCTGCTGCTGCTGCTGCTGCTGCTG 180
Db 121 TATTAACCTTACCCCAACCCCGTCTCTTAAACATGCTGCTGCTGCTGCTGCTGCTGCTG 180
Qy 181 ATGATTAAGGCGGTCAGAGATGTCTTTTAAACAGATGCTTGAAGGAGCATGTCTC 240
Db 181 ATGATTAAGGCGGTCAGAGATGTCTTTTAAACAGATGCTTGAAGGAGCATGTCTC 240
Qy 241 CTTAAGAGTCATCAACATCCCTATCTCAAGTACCCAGGACACAAAACCTGGGAAAG 300
Db 241 CTTAAGAGTCATCAACATCCCTATCTCAAGTACCCAGGACACAAAACCTGGGAAAG 300
Qy 301 CCGGAGGACCTCTGCTAGGAAAGCCAGATATTGTCCAGTTTCTCCCATGTGATAG 360
Db 301 CCGGAGGACCTCTGCTAGGAAAGCCAGATATTGTCCAGTTTCTCCCATGTGATAG 360
Qy 361 CTGGAATATGAGCTCTGGGAAAGAAACCTGACCGTCCCGACCCGACACCGCTA 420
Db 361 CTGGAATATGAGCTCTGGGAAAGAAACCTGACCGTCCCGACCCGACACCGCTA 420
Qy 421 AAGGCTCTGTGCTAGAGAGATTAAGTAAAGAGAGAGATGCTTGGAGTGAACA 480
Db 421 AAGGCTCTGTGCTAGAGAGATTAAGTAAAGAGAGAGATGCTTGGAGTGAACA 480
Qy 481 AGAGAAAGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 481 AGAGAAAGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Qy 541 GATTGTATGCTCCATCTTAAGTAAAGGAAAGAAACCCGCTTGAAGGAGTGGACCT 600
Db 541 GATTGTATGCTCCATCTTAAGTAAAGGAAAGAAACCCGCTTGAAGGAGTGGACCT 600
```


THIS PAGE BLANK (03/07)